Convergence of a finite difference scheme to weak solutions of the system of partial differential equation arising in mean field games

Yves Achdou, Alessio Porretta

CONVERGENCE OF A FINITE DIFFERENCE SCHEME TO WEAK SOLUTIONS OF THE SYSTEM OF PARTIAL DIFFERENTIAL EQUATION ARISING IN MEAN FIELD GAMES

YVES ACHDOU * AND ALESSIO PORRETTA †

Abstract. Mean field type models describing the limiting behavior of stochastic differential games as the number of players tends to $+\infty$, have been recently introduced by J-M. Lasry and P-L. Lions. Under suitable assumptions, they lead to a system of two coupled partial differential equations, a forward Bellman equation and a backward Fokker-Planck equations. Finite difference schemes for the approximation of such systems have been proposed in previous works. Here, we prove the convergence of these schemes towards a weak solution of the system of partial differential equations.

Key words. Mean field games, finite difference schemes, convergence.

AMS subject classifications. 65M06,65M012,9108,91A23,49L25

1. Introduction. Mean field type models describing the asymptotic behavior of stochastic differential games (Nash equilibria) as the number of players tends to $+\infty$ have recently been introduced by J-M. Lasry and P-L. Lions [22, 23, 24], and termed *mean field games* by the same authors. Related ideas have been developed independently in the engineering literature by Huang-Caines-Malhamé, see for example [20]. For brevity, the acronym *MFG* will sometimes be used for *mean field games*. Examples of MFG models with applications in economics and social sciences are proposed in [19, 2].

The simplest MFG model lead to systems of evolutive partial differential equations involving two unknown scalar functions: the density of the agents in a given state $x \in \mathbb{R}^d$, namely $m = m(t, x)$ and the value function $u = u(t, x)$. The present work is devoted to finite difference schemes for the systems of partial differential equations. Although the methods and the theoretical results obtained below can be easily generalized, the present work focuses on the two-dimensional case for the following reasons: 1) the one dimensional case is easier and allows too special arguments; 2) in dimension two, the description of the discrete methods discussed below remain fairly simple. Besides, several important applications of the mean field games theory are two-dimensional, in particular those related to crowd dynamics.

In the state-periodic setting, typical MFG model comprises the following system of partial differential equations in $(0, T) \times \mathbb{T}^2$

$$\frac{\partial u}{\partial t}(t, x) - \nu \Delta u(t, x) + H(x, \nabla u(t, x)) = F(m(t, x)), \tag{1.1}$$

$$\frac{\partial m}{\partial t}(t, x) + \nu \Delta m(t, x) + \operatorname{div}\left(m(t, x)\frac{\partial H}{\partial p}(x, \nabla u(t, x))\right) = 0, \tag{1.2}$$

with the initial and terminal conditions

$$u(0, x) = u_0(x), \quad m(T, x) = m_T(x), \quad \text{in } \mathbb{T}^2, \tag{1.3}$$

given a cost function u_0 and a probability density m_T.
Here, we denote by $\mathbb{T}^2 = [0, 1]^2$ the 2−dimensional unit torus, and by Δ, ∇ and

* Univ. Paris Diderot, Sorbonne Paris Cité, Laboratoire Jacques-Louis Lions, UMR 7598, UPMC, CNRS, F-75205 Paris, France. achdou@ljll-univ-paris-diderot.fr

† Università di Roma Tor Vergata, Dipartimento di Matematica, Via della Ricerca Scientifica 1, 00133 Roma, Italy

div, respectively, the Laplace, the gradient and the divergence operator acting on the state variable x. The parameter ν is the diffusion coefficient. Hereafter, we will always assume that $\nu > 0$. The system also involves the scalar Hamiltonian $H(x,p)$, which is assumed to be continuous, convex and $\mathcal{C}^1$ regular with respect to p. The notation $\frac{\partial H}{\partial p}(x,q)$ is used for the gradient of $p \mapsto H(x,p)$ at $p = q$. Finally, in the term $F(m(t,x))$, F is a continuous real valued function defined on $\mathbb{R}_+$. Hereafter the notation Q will be used for the space-time cylinder $(0,T) \times \mathbb{T}^2$.

We have chosen to focus on the case when the cost u_0 depends directly on x. In some realistic situations, the final cost may depend on the density of the players, i.e. $u_{|t=0} = \Phi_0[m_{|t=0}](x)$, where Φ_0 is an operator acting on probability densities, which may be local or not. We will not tackle this aspect, in order to keep the discussion as simple as possible. Similarly, by working on the torus $\mathbb{T}^2$, we avoid the discussion of the boundary conditions, but other boundary value problems with for example Dirichlet or Neumann conditions could be considered. It is also possible to consider different initial conditions than in (1.3): if there is a condition of the type $m(t=0,\cdot) = m_0$ instead of $u(t=0,\cdot) = u_0$, then the system models a planning problem, see [25] for a description of the model and mathematical results, and [26, 27] for new existence and uniqueness results.
System (1.1)-(1.2) consists then of a forward Bellman equation coupled with a backward Fokker-Planck equation. The forward-backward structure is an important feature of this system, which makes it necessary to design new strategies for its mathematical analysis (see [23, 24]) and for numerical approximation. The main results on the mathematical analysis of (1.1)-(1.2) are contained in the pioneering articles [23, 24], but many important aspects of the theory developed by J-M. Lasry and P-L. Lions on MFG are not published in journals or books. They can nevertheless be found in the videos of the lectures of P-L. Lions (in French) at Collège de France: see [25]. A very good introduction is also given in the notes by P. Cardaliaguet, [9], with a special emphasis on the deterministic case, i.e. $\nu = 0$ in (1.1)-(1.2). The survey of Gomes et al [16] also addresses interesting extensions of the model, and the so-called *master equation* first introduced in [25].

Depending on the data and on F and H, different notions of solutions can be relevant for (1.1)-(1.3): indeed, if the right hand side of (1.1) is replaced by $\Phi[m(t,\cdot)](x)$ where Φ is a nonlocal smoothing operator, mapping probability measures on $\mathbb{T}^2$ to $\mathcal{C}^1$ functions, if H depends smoothly on x and if the data u_0 and m_T are smooth, then classical solutions can be found, see [23, 24]. The same is true if e.g. H is Lipschitz continuous w.r.t. its second argument p and F in (1.1) is a continuous function. The situation is different in the case when H has a strictly superlinear growth with respect to p and F is a continuous function: in this case, one has to look for weak solutions, see [24] and the recent article [28] which is devoted to weak solutions to Fokker-Planck equations and to the system (1.1)-(1.2).
Since the (semi-)analytic solutions of the MFG system do not exist in general, any attempt to apply MFG models and to get qualitative/quantitative information from them must rely on numerical simulations and scientific computing. Therefore, the research has also been active on numerical methods for approximating (1.1)-(1.3): a numerical method based on the reformulation of the model as an optimal control problem for the Fokker-Planck equation with an application in economics was proposed in [21]. Discrete time, finite state space mean field games were discussed in [15]. We also refer to [17, 18] for a specific constructive approach when the Hamiltonian is quadratic. Finally, semi-Lagrangian approximations have been studied in [11, 10].
The finite difference method described and studied below has first been proposed and discussed in [5, 3]. It will be reviewed in § 2. The numerical scheme basically

relies on monotone approximations of the Hamiltonian and on a suitable weak formulation of the Fokker-Planck equation. It has several important features:

- existence and uniqueness for the discretized problems can be obtained by similar arguments as those used in the continuous case,
- they are robust when $\nu \to 0$ (the deterministic limit of the models),
- bounds on the solutions, which are uniform in the grid step, can be proved under reasonable assumptions on the data.

A first result on the convergence to classical solutions was given in [5]. The issue of convergence was studied with more details in [4, 1]: in these works, the starting point/assumption was the existence of a classical solution of (1.1)-(1.3). The proof of convergence mainly consisted in plugging the classical solution into the system of equations arising from the finite difference method, and use the consistency and stability properties of the scheme in order to get estimates and pass to the limit.
In the present work, the goal is different: we wish to prove that as the grid steps tend to zero, the solution of the discretized MFG system converges to a weak solution of (1.1)-(1.3), without assuming the existence of the latter; so this work will supply as a by-product a new strategy for proving the existence of weak solutions. One key step will be to obtain a priori estimates on the solutions to the discrete systems, and these will mainly come from the fact that the structure of the MFG system is preserved by the chosen finite difference method. This step will be achieved in § 4. Note that § 4.3 is concerned with a priori estimates for the discrete version of the Fokker-Planck equation (1.2). These estimates may have their own interest, independently from MFG models, and may be put in relation with recent works of Gallouët et al [14] in the context of finite volume methods. Once these estimates are obtained, the most important difficulty will be to pass to the limit in the discrete Bellman equation. The strategy to that purpose is to first prove some L^1 compactness of the sequence of state-gradients of the discrete solutions, then to adapt some techniques proposed by Boccardo, Murat and Puel, see [8], for studying weak solutions of (1.1). This will done in § 5.1 and 6. Here also, we think that the passage to the limit in the discrete Bellman equation may have an interest for itself.

2. Finite difference schemes. In the present paragraph, we discuss the finite difference method originally proposed in [5].
Let N_T be a positive integer and $\Delta t = T/N_T$, $t_n = n\Delta t$, $n = 0, \dots, N_T$. Let $\mathbb{T}_h^2$ be a uniform grid on the torus with mesh step h, (assuming that $1/h$ is an integer N_h), and x_{ij} denote a generic point in $\mathbb{T}_h^2$. The values of u and m at $(x_{i,j}, t_n)$ are respectively approximated by $u_{i,j}^n$ and $m_{i,j}^n$. Let u^n (resp. m^n) be the vector containing the values $u_{i,j}^n$ (resp. $m_{i,j}^n$), for $0 \le i, j < N_h$ indexed in the lexicographic order. Hereafter, such vectors will be termed *grid functions on* $\mathbb{T}_h^2$ or simply grid functions. For all grid functions z, all i and j, we agree that $z_{i,j} = z_{(i \bmod N_h),(j \bmod N_h)}$.

Elementary finite difference operators. Let us introduce the elementary finite difference operators

$$(D_1^+ u)_{i,j} = \frac{u_{i+1,j} - u_{i,j}}{h} \quad \text{and} \quad (D_2^+ u)_{i,j} = \frac{u_{i,j+1} - u_{i,j}}{h}, \tag{2.1}$$

and define $D_h u$ as the grid function with values in $\mathbb{R}^2$:

$$(D_h u)_{i,j} = \Big((D_1^+ u)_{i,j}, (D_2^+ u)_{i,j} \Big) \in \mathbb{R}^2. \tag{2.2}$$

Let $[\nabla_h u]_{i,j}$ be the collection of the four possible one sided finite differences at $x_{i,j}$:

$$[\nabla_h u]_{i,j} = \Big((D_1^+ u)_{i,j}, (D_1^+ u)_{i-1,j}, (D_2^+ u)_{i,j}, (D_2^+ u)_{i,j-1} \Big) \in \mathbb{R}^4. \tag{2.3}$$

We will also need the standard five point discrete Laplace operator

$$(\Delta_h u)_{i,j} = -\frac{1}{h^2}(4u_{i,j} - u_{i+1,j} - u_{i-1,j} - u_{i,j+1} - u_{i,j-1}).$$

For a set $v = (v^n)_{n=0,\dots,N_T}$, where v^n is grid functions on $\mathbb{T}_h^2$, it will be convenient to define the family of grid functions:

$$\partial_{t,\Delta t} v \equiv \left(\frac{v^{n+1} - v^n}{\Delta t}\right)_{n=0,\dots,N_T-1}. \tag{2.4}$$

Numerical Hamiltonian. In order to approximate the term $H(x, \nabla u)$ in (1.1), we consider a numerical Hamiltonian $g : \mathbb{T}^2 \times \mathbb{R}^4 \to \mathbb{R}$, $(x, q_1, q_2, q_3, q_4) \mapsto g\,(x, q_1, q_2, q_3, q_4)$. Hereafter we will often assume that the following conditions hold:

($\mathbf{g_1}$) *monotonicity*: g is nonincreasing with respect to q_1 and q_3 and nondecreasing with respect to q_2 and q_4.

($\mathbf{g_2}$) *consistency:* $g\,(x, q_1, q_1, q_2, q_2) = H(x, q), \quad \forall x \in \mathbb{T}^2, \forall q = (q_1, q_2) \in \mathbb{R}^2$.

($\mathbf{g_3}$) *regularity*: g is continuous and of class $\mathcal{C}^1$ w.r.t. (q_1, q_2, q_3, q_4).

($\mathbf{g_4}$) *convexity* : $(q_1, q_2, q_3, q_4) \mapsto g\,(x, q_1, q_2, q_3, q_4)$ is convex.

We will approximate $H(\cdot, \nabla u)(x_{i,j})$ by $g(x_{i,j}, [\nabla_h u]_{i,j})$.
Standard examples of numerical Hamiltonians fulfilling these requirements are provided by Lax-Friedrichs or upwind schemes, see [5]. For Hamiltonians of the form $H(x, p) = \mathcal{H}(x) + |p|^\beta$, $\beta \in (1, \infty)$, we may choose

$$g(x, q) = \mathcal{H}(x) + G(q_1^-, q_2^+, q_3^-, q_4^+), \tag{2.5}$$

where, for a real number r, $r^+ = \max(r, 0)$ and $r^- = \max(-r, 0)$ and where $G : (\mathbb{R}_+)^4 \to \mathbb{R}_+$ is given by

$$G(p) = |p|^\beta = (p_1^2 + p_2^2 + p_3^2 + p_4^2)^{\frac{\beta}{2}}. \tag{2.6}$$

Discrete Bellman equation. The discrete version of the Bellman equation is obtained by applying a semi-implicit Euler scheme to (1.1),

$$\frac{u_{i,j}^{n+1} - u_{i,j}^n}{\Delta t} - \nu(\Delta_h u^{n+1})_{i,j} + g(x_{i,j}, \left[\nabla_h u^{n+1}\right]_{i,j}) = F(m_{i,j}^n), \tag{2.7}$$

for all points in $\mathbb{T}_h^2$ and all n, $0 \le n < N_T$, where all the discrete operators have been introduced above. Given $(m^n)_{n=0,\dots,N_T-1}$, (2.7) and the initial condition $u_{i,j}^0 = u_0(x_{i,j})$ for all (i, j) completely characterizes $(u^n)_{0 \le n \le N_T}$.

Discrete Fokker-Planck equation. In order to approximate equation (1.2), it is convenient to consider its weak formulation which involves in particular the term

$$\int_{\mathbb{T}^2} \mathrm{div}\left(m \frac{\partial H}{\partial p}(x, \nabla u)\right) w(x)\, dx.$$

By periodicity,

$$\int_{\mathbb{T}^2} \mathrm{div}\left(m \frac{\partial H}{\partial p}(x, \nabla u)\right) w(x)\, dx = -\int_{\mathbb{T}^2} m(x) \frac{\partial H}{\partial p}(x, \nabla u(x)) \cdot \nabla w(x)\, dx$$

holds for any test function w. The right hand side in the identity above will be approximated by

$$-h^2 \sum_{i,j} m_{i,j} \nabla_q g(x_{i,j}, [\nabla_h u]_{i,j}) \cdot [\nabla_h w]_{i,j} = h^2 \sum_{i,j} \mathcal{T}_{i,j}(u, m) w_{i,j},$$

where the transport operator $\mathcal{T}$ is defined as follows:

$$\mathcal{T}_{i,j}(u,m) = \frac{1}{h}\left(\begin{array}{l}\left(\begin{array}{l} m_{i,j}\dfrac{\partial g}{\partial q_1}(x_{i,j},[\nabla_h u]_{i,j}) - m_{i-1,j}\dfrac{\partial g}{\partial q_1}(x_{i-1,j},[\nabla_h u]_{i-1,j}) \\ +m_{i+1,j}\dfrac{\partial g}{\partial q_2}(x_{i+1,j},[\nabla_h u]_{i+1,j}) - m_{i,j}\dfrac{\partial g}{\partial q_2}(x_{i,j},[\nabla_h u]_{i,j}) \end{array}\right) \\ + \\ \left(\begin{array}{l} m_{i,j}\dfrac{\partial g}{\partial q_3}(x_{i,j},[\nabla_h u]_{i,j}) - m_{i,j-1}\dfrac{\partial g}{\partial q_3}(x_{i,j-1},[\nabla_h u]_{i,j-1}) \\ +m_{i,j+1}\dfrac{\partial g}{\partial q_4}(x_{i,j+1},[\nabla_h u]_{i,j+1}) - m_{i,j}\dfrac{\partial g}{\partial q_4}(x_{i,j},[\nabla_h u]_{i,j}) \end{array}\right)\end{array}\right). \tag{2.8}$$

The discrete version of equation (1.2) is chosen as follows:

$$\frac{m_{i,j}^{n+1} - m_{i,j}^n}{\Delta t} + \nu(\Delta_h m^n)_{i,j} + \mathcal{T}_{i,j}(u^{n+1}, m^n) = 0, \tag{2.9}$$

for all $n = 0, \dots, N_T - 1$. This scheme is implicit w.r.t. to m and explicit w.r.t. u because the considered Fokker-Planck equation is backward. Given u this is a system of linear equations for m. We introduce the compact and convex set

$$\mathcal{K}_h = \{(m_{i,j})_{0\le i,j<N_h} : h^2 \sum_{i,j} m_{i,j} = 1; \quad m_{i,j} \ge 0\} \tag{2.10}$$

which can be viewed as the set of the discrete probability measures. It is easy to see that if m^n satisfies (2.9) for $0 \le n < N_T$ and if $m^{N_T} \in \mathcal{K}_h$, then $m^n \in \mathcal{K}_h$ for all n, $0 \le n < N_T$.

Remark 1. *An important property of $\mathcal{T}$ is that the operator $m \mapsto \big(-\nu(\Delta_h m)_{i,j} - \mathcal{T}_{i,j}(u,m)\big)_{i,j}$ is the adjoint of the linearized version of the operator $u \mapsto \big(-\nu(\Delta_h u)_{i,j} + g(x_{i,j}, [\nabla_h u]_{i,j})\big)_{i,j}$.*
This property implies that the structure of (1.1)-(1.2) is preserved in the discrete version (2.7)-(2.9). In particular, it implies the uniqueness result stated in Theorem 2.2 below.

Summary. The fully discrete scheme for system (1.1),(1.2),(1.3) is therefore the following: for all $0 \le i,j < N_h$ and $0 \le k < N_T$

$$\left\{\begin{array}{ll} \frac{u_{i,j}^{k+1} - u_{i,j}^k}{\Delta t} - \nu(\Delta_h u^{k+1})_{i,j} + g(x_{i,j}, \left[\nabla_h u^{k+1}\right]_{i,j}) & = F(m_{i,j}^k), \\ \frac{m_{i,j}^{k+1} - m_{i,j}^k}{\Delta t} + \nu(\Delta_h m^k)_{i,j} + \mathcal{T}_{i,j}(u^{k+1}, m^k) & = 0, \end{array}\right. \tag{2.11}$$

with the initial and terminal conditions

$$u_{i,j}^0 = u_0(x_{i,j}), \qquad m_{i,j}^{N_T} = \frac{1}{h^2}\int_{|x-x_{i,j}|_\infty \le h/2} m_T(x)dx, \qquad 0 \le i,j < N_h. \tag{2.12}$$

The following theorem was proved in [5] (using essentially Brouwer's fixed point theorem and estimates on the solutions of the discrete Bellman equation):

THEOREM 2.1. *Assume that $(\mathbf{g_1})$–$(\mathbf{g_3})$ hold, that u_0 is continuous on $\mathbb{T}^2$ and that $m_T \in L^1(\mathbb{T}^2)$ is a probability density, i.e. $m_T \ge 0$ and $\int_{\mathbb{T}^2} m_T(x)dx = 1$; then (2.11)–(2.12) has a solution such that $m^n \in \mathcal{K}_h$, $\forall n$.*

Since (2.11)-(2.12) has exactly the same structure as the continuous problem (1.1)-(1.3), uniqueness has been obtained in [5] with the same arguments as in [23]:

THEOREM 2.2. *Assume that $(\mathbf{g_1})$–$(\mathbf{g_4})$ hold and that F is nondecreasing then (2.11)–(2.12) has a unique solution.*

Remark 2. *Efficient algorithms for solving system (2.11)-(2.12) require special efforts, essentially because of the forward-backward structure already discussed above. We refer to [5] for the description of possible algorithms and numerical results.*

3. Running assumptions and statement of the main result. We now summarize the assumptions that will be made in the whole work.

- u_0 is a continuous function on $\mathbb{T}^2$
- m_T is a nonnegative function in $L^\infty(\mathbb{T}^2)$ such that $\int_{\mathbb{T}^2} m_T(x)dx = 1$
- F is a continuous function on $\mathbb{R}^+$, which is bounded from below.
- The Hamiltonian[1] $(x,p) \mapsto H(x,p)$ is assumed to be convex with respect to p and $\mathcal{C}^1$ regular w.r.t. x and p.
- The discrete Hamiltonian g satisfies $(\mathbf{g_1})$-$(\mathbf{g_4})$ and the further assumption
  $(\mathbf{g_5})$ There exist positive constants c_1, c_2, c_3, c_4 such that

$$g_q(x,q)\cdot q - g(x,q) \geq c_1|g_q(x,q)|^2 - c_2, \tag{3.1}$$

$$|g_q(x,q)| \leq c_3|q| + c_4. \tag{3.2}$$

Take for example g as in (2.5) (2.6). It is clear that $g_q(x,q)\cdot q = \beta G(q_1^-, q_2^+, q_3^-, q_4^+)$, hence $g_q(x,q)\cdot q - g(x,q) = (\beta-1)G(q_1^-, q_2^+, q_3^-, q_4^+) - \mathcal{H}(x)$. Since $|g_q(x,q)|^2 = \beta^2 \left(G(q_1^-, q_2^+, q_3^-, q_4^+)\right)^{2\frac{\beta-1}{\beta}}$, we see that $(\mathbf{g_5})$ holds if $1 \leq \beta \leq 2$.

We can now state the main result of this article, which establishes the convergence of the solutions of the finite difference scheme towards a weak solution of the continuous mean field games system.

THEOREM 3.1. *Let $(u^n), (m^n)$ be a solution of the discrete system (2.11)-(2.12) and $u_{h,\Delta t}$, $m_{h,\Delta t}$ be the piecewise constant functions which take the values $u_{i,j}^{n+1}$ and $m_{i,j}^n$, respectively, in $(t_n, t_{n+1}) \times (ih - h/2, ih + h/2) \times (jh - h/2, jh + h/2)$. There exists a subsequence of h and Δt (not relabeled) and functions $\tilde{u}$, $\tilde{m}$, which belong to $L^\alpha(0,T; W^{1,\alpha}(\mathbb{T}^2))$ for any $\alpha \in [1, \frac{4}{3})$, such that $u_{h,\Delta t} \to \tilde{u}$ and $m_{h,\Delta t} \to \tilde{m}$ in $L^\beta(Q)$ for all $\beta \in [1,2)$, and $(\tilde{u}, \tilde{m})$ is a weak solution to the system (1.1)-(1.3) in the following sense:*

(i) $H(\cdot, D\tilde{u}) \in L^1(Q)$, $\tilde{m}F(\tilde{m}) \in L^1(Q)$, $\tilde{m}[H_p(\cdot, D\tilde{u})\cdot D\tilde{u} - H(\cdot, D\tilde{u})] \in L^1(Q)$
(ii) $(\tilde{u}, \tilde{m})$ *satisfies (1.1)-(1.2) in the sense of distributions*
(iii) $\tilde{u}, \tilde{m} \in C^0([0,T]; L^1(\mathbb{T}^2))$ *and* $\tilde{u}|_{t=0} = u_0$, $\tilde{m}|_{t=T} = m_T$.

Remark 3. *We recall that, if F is nondecreasing and $p \mapsto H(x,p)$ is strictly convex at infinity, it is proved in [28] that weak solutions are unique whenever H satisfies the structure conditions*

$$\begin{aligned} H_p(t,x,p)\cdot p &\geq r\, H(t,x,p) - \gamma \\ |H_p(t,x,p)| &\leq \beta\,(1 + |p|^{r-1}) \\ H(t,x,p) &\geq \alpha|p|^r - \gamma \end{aligned}$$

for some $r \in (1,2]$ and some positive constant α, β, γ.

Therefore, in this case the convergence established in the above theorem holds for the whole sequence, and not only for a subsequence.

4. A priori estimates.

4.1. Norms and semi-norms.

It is useful to define the following norms and semi-norms:

[1] il carattere C^1 rispetto a x dove viene realmente usato ? Idem per la g

for a grid function $v \equiv (v_{i,j})_{i,j}$, we define

$$\|v\|_{L^s(\mathbb{T}_h^2)} = \left(h^2 \sum_{i,j} |v_{i,j}|^s \right)^{\frac{1}{s}}, \tag{4.1}$$

$$|v|_{W^{1,s}(\mathbb{T}_h^2)} = \left(h^2 \sum_{i,j} \left((D_1^+ v_{i,j})^2 + (D_2^+ v_{i,j})^2 \right)^{\frac{s}{2}} \right)^{\frac{1}{s}} \tag{4.2}$$

$$\|v\|_{W^{1,s}(\mathbb{T}_h^2)} = \left(\|v\|^s_{L^s(\mathbb{T}_h^2)} + |v|^s_{W^{1,s}(\mathbb{T}_h^2)} \right)^{\frac{1}{s}}, \tag{4.3}$$

where $D_1^+ v$ and $D_2^+ v$ are defined in (2.1). We shall also write $|v|_{H^1(\mathbb{T}_h^2)} = |v|_{W^{1,2}(\mathbb{T}_h^2)}$, $\|v\|_{H^1(\mathbb{T}_h^2)} = \|v\|_{W^{1,2}(\mathbb{T}_h^2)}$, and define the discrete L^2 scalar product:

$$(v, w)_{L^2(\mathbb{T}_h^2)} = h^2 \sum_{i,j} v_{i,j} w_{i,j}.$$

We recall the discrete Sobolev inequality: for any $s < \infty$, there exists a constant C such that for any grid function v,

$$\|v\|_{L^s(\mathbb{T}_h^2)} \le C \left(\|v\|_{L^2(\mathbb{T}_h^2)} + |v|_{H^1(\mathbb{T}_h^2)} \right).$$

For $s > 1$, we define the dual norm $\|v\|_{W^{-1,s'}(\mathbb{T}_h^2)}$, $\frac{1}{s} + \frac{1}{s'} = 1$ by

$$\|v\|_{W^{-1,s'}(\mathbb{T}_h^2)} = \sup_{w \neq 0} \frac{(v, w)_{L^2(\mathbb{T}_h^2)}}{\|w\|_{W^{1,s}(\mathbb{T}_h^2)}}.$$

Define $Q_{h,\Delta t} = \Delta t \{0, \ldots, N_T - 1\} \times \mathbb{T}_h^2$. For a function w defined on $Q_{h,\Delta t}$, $w \equiv (w^n_{i,j})_{i,j}$, $0 \le n \le N_T$, we define for $s \in [1, +\infty)$,

$$\|w\|_{L^s(Q_{h,\Delta t})} = \left(\Delta t \sum_{n=0}^{N_T} \|w^n\|^s_{L^s(\mathbb{T}_h^2)} \right)^{\frac{1}{s}}. \tag{4.4}$$

4.2. First estimates. Hereafter, the constants appearing in the a priori estimates, for example c, C, are independent of h and Δt. In this paragraph, we state the first a priori estimates stemming from the structure of the system. Although we have already given the set of running assumptions, we think that it may be useful to specify which assumptions are really required by each particular result.

LEMMA 4.1. *Under Assumptions* ($\mathbf{g_1}$) *and* ($\mathbf{g_3}$), *if F is bounded from below by a constant $\underline{F}$, u_0 is continuous on $\mathbb{T}^2$, then for all i, j, n,*

$$u^n_{i,j} \ge \underline{u} - T \left(\underline{F} - \max_{x \in \mathbb{T}^2} H(x, 0) \right)^-, \tag{4.5}$$

where $\underline{u} = \min_{x \in \mathbb{T}^2} u_0(x)$.

LEMMA 4.2. *Under Assumptions* ($\mathbf{g_1}$), ($\mathbf{g_3}$) *and* ($\mathbf{g_5}$), *if F is bounded from below by $\underline{F}$, u_0 is continuous on $\mathbb{T}^2$ and m_T is bounded from above by $\bar{m}_T$, then there exists a constant C such that*

$$h^2 \Delta t \sum_{k=0}^{N_T - 1} \sum_{i,j} m^k_{i,j} \left| g_q(x_{i,j}, [\nabla_h u^{k+1}]_{i,j}) \right|^2 \le C, \tag{4.6}$$

$$h^2 \Delta t \sum_{k=0}^{N_T - 1} \sum_{i,j} g(x_{i,j}, [\nabla_h u^{k+1}]_{i,j}) \le C, \tag{4.7}$$

$$h^2 \Delta t \sum_{k=0}^{N_T - 1} \sum_{i,j} m^k_{i,j} F(m^k_{i,j}) \le C. \tag{4.8}$$

Proof. Consider $\tilde{u}^n_{i,j} = n\Delta t F(\bar{m}_T)$ for all i,j,n. We get immediately

$$\frac{\tilde{u}^{n+1}_{i,j} - \tilde{u}^n_{i,j}}{\Delta t} - \nu(\Delta_h \tilde{u}^{n+1})_{i,j} = F(\bar{m}_T) \tag{4.9}$$

Subtract (4.9) from (2.7) and multiply the resulting equation by $m^n_{i,j} - \bar{m}_T$. Similarly, multiply (2.9) by $u^{n+1}_{i,j} - \tilde{u}^{n+1}_{i,j}$. Adding the two resulting identities and summing with respect to n, one gets:

$$\begin{aligned}
&h^2\Delta t \sum_{k=0}^{N_T-1}\sum_{i,j} m^k_{i,j}\left(g_q(x_{i,j},\left[\nabla_h u^{k+1}\right]_{i,j})\cdot\left[\nabla_h u^{k+1}\right]_{i,j} - g(x_{i,j},\left[\nabla_h u^{k+1}\right]_{i,j})\right)\\
&+h^2\Delta t \sum_{k=0}^{N_T-1}\sum_{i,j} \bar{m}_T g(x_{i,j},\left[\nabla_h u^{k+1}\right]_{i,j})\\
&+h^2\Delta t \sum_{k=0}^{N_T-1}\sum_{i,j}(m^k_{i,j}-\bar{m}_T)(F(m^k_{i,j})-F(\bar{m}_T))\\
=&(m^{N_T}-\bar{m}_T, u^{N_T}-T\,F(\bar{m}_T))_{L^2(\mathbb{T}_h^2)} - (m^0-\bar{m}_T,u^0)_{L^2(\mathbb{T}_h^2)}.
\end{aligned} \tag{4.10}$$

1. Since $m^{N_T}-\bar{m}_T$ is nonpositive with a bounded mass, and since u^n is bounded from below, see (4.5), the term $(m^{N_T}-\bar{m}_T, u^{N_T}-T\,F(\bar{m}_T))_{L^2(\mathbb{T}_h^2)}$ in the right hand side of (4.10) is bounded from above by a constant independent of h and Δt.
2. It is straightforward to see that $(m^0-\bar{m}_T,u^0)_{L^2(\mathbb{T}_h^2)} \le (1+\bar{m}_T)\|u_0\|_\infty$.
3. Since F is continuous, there exists a constant c such that $F(t)\le \frac{1}{2\bar{m}_T}tF(t)+c$, $\forall t\ge 0$. Hence,
$$h^2\Delta t\sum_{k=0}^{N_T-1}\sum_{i,j}(m^k_{i,j}-\bar{m}_T)F(m^k_{i,j}) \ge \frac{1}{2}h^2\Delta t\sum_{k=0}^{N_T-1}\sum_{i,j} m^k_{i,j}F(m^k_{i,j}) - c.$$
4. Finally, $h^2\Delta t\sum_{k=0}^{N_T-1}\sum_{i,j}(m^k_{i,j}-\bar{m}_T)F(\bar{m}_T) = T(1-\bar{m}_T)F(\bar{m}_T)$

From these observations, (4.6), (4.7) and (4.8) follow from (4.10) and (3.1). ∎

4.3. A priori estimates from the discrete Fokker-Planck equation. The following estimates for the Fokker-Planck equation may have their own interest:

LEMMA 4.3. *Assume* ($\mathbf{g_1}$) *and* ($\mathbf{g_3}$). *Let ψ be a non decreasing and concave function defined on $\mathbb{R}_+$. For any grid function $v=(v_{i,j})$, any positive grid function $m=(m_{i,j})$ and any positive number η,*

$$\begin{aligned}
&\sum_{i,j}\mathcal{T}_{i,j}(v,m)\psi(m_{i,j}) \le \frac{\eta}{2}\sum_{i,j}[\nabla_h\psi(m)]_{i,j}\cdot[\nabla_h m]_{i,j}\\
&+\frac{1}{2\eta}\sum_{i,j} m^2_{i,j}\psi'(m_{i,j})\left(\frac{\partial g}{\partial q_1}(x_{i,j},[\nabla_h v]_{i,j})\right)^2 1_{\{m_{i+1,j}>m_{i,j}\}}\\
&+\frac{1}{2\eta}\sum_{i,j} m^2_{i,j}\psi'(m_{i,j})\left(\frac{\partial g}{\partial q_2}(x_{i,j},[\nabla_h v]_{i,j})\right)^2 1_{\{m_{i-1,j}>m_{i,j}\}}\\
&+\frac{1}{2\eta}\sum_{i,j} m^2_{i,j}\psi'(m_{i,j})\left(\frac{\partial g}{\partial q_3}(x_{i,j},[\nabla_h v]_{i,j})\right)^2 1_{\{m_{i,j+1}>m_{i,j}\}}\\
&+\frac{1}{2\eta}\sum_{i,j} m^2_{i,j}\psi'(m_{i,j})\left(\frac{\partial g}{\partial q_4}(x_{i,j},[\nabla_h v]_{i,j})\right)^2 1_{\{m_{i,j-1}>m_{i,j}\}}.
\end{aligned} \tag{4.11}$$

In particular, if m does not vanish, then for $\psi(z) = \ln(z)$,

$$\begin{aligned} & \sum_{i,j} \mathcal{T}_{i,j}(v,m)\ln(m_{i,j}) \\ \leq \quad & \frac{\eta}{2}\sum_{i,j} [\nabla_h \ln(m)]_{i,j} \cdot [\nabla_h m]_{i,j} + \frac{1}{2\eta}\sum_{i,j} m_{i,j} \left| g_q(x_{i,j}, [\nabla_h v]_{i,j}) \right|^2. \end{aligned} \tag{4.12}$$

Proof. By the definition of $\mathcal{T}$, we can split the sum $S = \sum_{i,j} \mathcal{T}_{i,j}(v,m)\psi(m_{i,j})$ as follows:

$$S = -\sum_{i,j} m_{i,j} \nabla_q g(x_{i,j}, [\nabla_h v]_{i,j}) \cdot [\nabla_h \psi(m)]_{i,j} = S_1 + S_2 + S_3 + S_4,$$

where

$$\begin{aligned} S_1 &= -\frac{1}{h}\sum_{i,j} m_{i,j} \frac{\partial g}{\partial q_1}(x_{i,j}, [\nabla_h v]_{i,j})(\psi(m_{i+1,j}) - \psi(m_{i,j})), \\ S_2 &= -\frac{1}{h}\sum_{i,j} m_{i,j} \frac{\partial g}{\partial q_2}(x_{i,j}, [\nabla_h v]_{i,j})(\psi(m_{i,j}) - \psi(m_{i-1,j})), \\ S_3 &= -\frac{1}{h}\sum_{i,j} m_{i,j} \frac{\partial g}{\partial q_3}(x_{i,j}, [\nabla_h v]_{i,j})(\psi(m_{i,j+1}) - \psi(m_{i,j})), \\ S_4 &= -\frac{1}{h}\sum_{i,j} m_{i,j} \frac{\partial g}{\partial q_4}(x_{i,j}, [\nabla_h v]_{i,j})(\psi(m_{i,j}) - \psi(m_{i,j-1})). \end{aligned}$$

It is enough to focus on S_1 since the same arguments can be used for the other sums. Since g is nonincreasing w.r.t. q_1,

$$S_1 \leq -\frac{1}{h}\sum_{i,j} m_{i,j} \frac{\partial g}{\partial q_1}(x_{i,j}, [\nabla_h v]_{i,j})(\psi(m_{i+1,j}) - \psi(m_{i,j}))_+.$$

Since ψ is nondecreasing, if $m_{i+1,j} > m_{i,j}$, the factor $(\psi(m_{i+1,j}) - \psi(m_{i,j}))_+$ can be rewritten

$$\left(\frac{\psi(m_{i+1,j}) - \psi(m_{i,j})}{m_{i+1,j} - m_{i,j}}\right)^{\frac{1}{2}} \Big((\psi(m_{i+1,j}) - \psi(m_{i,j}))(m_{i+1,j} - m_{i,j})\Big)^{\frac{1}{2}}.$$

Since ψ is nondecreasing and concave, $m_{i+1,j} > m_{i,j}$ implies that

$$0 \leq \frac{\psi(m_{i+1,j}) - \psi(m_{i,j})}{m_{i+1,j} - m_{i,j}} \leq \psi'(m_{i,j}).$$

Hence, if $m_{i+1,j} \geq m_{i,j}$, then

$$(\psi(m_{i+1,j}) - \psi(m_{i,j}))_+ \leq \Big(\psi'(m_{i,j})(\psi(m_{i+1,j}) - \psi(m_{i,j}))(m_{i+1,j} - m_{i,j})\Big)^{\frac{1}{2}},$$

which implies that

$$\begin{aligned} S_1 \quad \leq \quad & \frac{1}{2\eta}\sum_{i,j} m_{i,j}^2 \psi'(m_{i,j}) \left(\frac{\partial g}{\partial q_1}(x_{i,j}, [\nabla_h v]_{i,j})\right)^2 1_{\{m_{i+1,j} > m_{i,j}\}} \\ & + \frac{\eta}{2h^2}\sum_{i,j} (\psi(m_{i+1,j}) - \psi(m_{i,j}))(m_{i+1,j} - m_{i,j}) 1_{\{m_{i+1,j} > m_{i,j}\}}. \end{aligned}$$

$\square$

LEMMA 4.4. *Assume* ($\mathbf{g_1}$) *and* ($\mathbf{g_3}$). *Let ψ be a non decreasing and concave function defined on $\mathbb{R}_+$. For any positive grid functions $(m_{i,j}^k)$, $k = 0,\dots,N_T$,*

$$\sum_{k=n+1}^{N_T}\sum_{i,j} m_{i,j}^k(\psi(m_{i,j}^k) - \psi(m_{i,j}^{k-1})) \geq -\sum_{k=n+1}^{N_T}\sum_{i,j} m_{i,j}^k \psi'(m_{i,j}^k)(m_{i,j}^{k-1} - m_{i,j}^k)_+. \tag{4.13}$$

If $m^k \in \mathcal{K}_h$ for all $k \in \{0,\dots,N_T\}$ and does not vanish, then

$$h^2\sum_{k=n}^{N_T-1}\sum_{i,j}\left(m_{i,j}^{k+1} - m_{i,j}^k\right)\ln(m_{i,j}^k) \leq h^2\sum_{i,j} m_{i,j}^{N_T}\ln(m_{i,j}^{N_T}) - h^2\sum_{i,j} m_{i,j}^n\ln(m_{i,j}^n) + 1. \tag{4.14}$$

Proof. Since ψ is non decreasing,

$$\sum_{k=n+1}^{N_T}\sum_{i,j} m_{i,j}^k(\psi(m_{i,j}^k) - \psi(m_{i,j}^{k-1})) \geq \sum_{k=n+1}^{N_T}\sum_{i,j} m_{i,j}^k(\psi(m_{i,j}^k) - \psi(m_{i,j}^{k-1}))1_{\{m_{i,j}^k < m_{i,j}^{k-1}\}}.$$

From the concavity of ψ, if $m_{i,j}^k < m_{i,j}^{k-1}$, then $\psi(m_{i,j}^k) - \psi(m_{i,j}^{k-1}) \geq \psi'(m_{i,j}^k)(m_{i,j}^k - m_{i,j}^{k-1})$. Then (4.13) follows from the last two points.
Let us turn to (4.14): for any $\epsilon > 0$,

$$\begin{aligned}h^2\sum_{k=n}^{N_T-1}\sum_{i,j}\left(m_{i,j}^{k+1} - m_{i,j}^k\right)\ln(m_{i,j}^k + \epsilon) =& h^2\sum_{i,j} m_{i,j}^{N_T}\ln(m_{i,j}^{N_T} + \epsilon) - h^2\sum_{i,j} m_{i,j}^n\ln(m_{i,j}^n + \epsilon)\\ &- h^2\sum_{k=n+1}^{N_T}\sum_{i,j} m_{i,j}^k(\ln(m_{i,j}^k + \epsilon) - \ln(m_{i,j}^{k-1} + \epsilon)),\end{aligned}$$

and (4.13) with $\psi(z) = \ln(z + \epsilon)$ yields

$$\begin{aligned}-h^2\sum_{k=n+1}^{N_T}\sum_{i,j} m_{i,j}^k(\ln(m_{i,j}^k + \epsilon) - \ln(m_{i,j}^{k-1} + \epsilon)) &\leq h^2\sum_{k=n+1}^{N_T}\sum_{i,j}\frac{m_{i,j}^k}{m_{i,j}^k + \epsilon}(m_{i,j}^{k-1} - m_{i,j}^k)_+\\ &\leq h^2\sum_{k=n+1}^{N_T}\sum_{i,j}(m_{i,j}^{k-1} - m_{i,j}^k)_+ \leq 1,\end{aligned}$$

where the last estimate comes from the fact that the grid functions m^k all belong to $\mathcal{K}_h$. Hence,

$$h^2\sum_{k=n}^{N_T-1}\sum_{i,j}\left(m_{i,j}^{k+1} - m_{i,j}^k\right)\ln(m_{i,j}^k + \epsilon) \leq h^2\sum_{i,j} m_{i,j}^{N_T}\ln(m_{i,j}^{N_T} + \epsilon) - h^2\sum_{i,j} m_{i,j}^n\ln(m_{i,j}^n + \epsilon) + 1.$$

and (4.14) is obtained by letting ϵ tend to 0. ∎

LEMMA 4.5. *If $m^{N_T} \in \mathcal{K}_h$ and* ($\mathbf{g_1}$) ($\mathbf{g_3}$) *hold, then there exists a constant C such that, for any number η, $0 < \eta < \nu$, a solution $(m_{i,j}^n)$ of (2.9) satisfies*

$$\begin{aligned}&\max_n h^2\sum_{i,j} m_{i,j}^n|\ln(m_{i,j}^n)| + (\nu - \eta)\Delta t\sum_{k=0}^{N_T-1}\left|\sqrt{m^k}\right|_{H^1(\mathbb{T}_h^2)}^2\\ \leq\ & C + h^2\sum_{i,j} m_{i,j}^{N_T}|\ln(m_{i,j}^{N_T})| + \frac{h^2\Delta t}{2\eta}\sum_{k=0}^{N_T-1}\sum_{i,j} m_{i,j}^k\left|g_q\left(x_{i,j}, \left[\nabla_h u^{k+1}\right]_{i,j}\right)\right|^2.\end{aligned} \tag{4.15}$$

For all $\alpha \in [1,2)$, there exists a constant c such that

$$
\begin{aligned}
&\|m\|^{\alpha}_{L^{\alpha}(Q_{h,\Delta t})}\\
\le & c\left(1 + h^2 \sum_{i,j} m_{i,j}^{N_T} |\ln(m_{i,j}^{N_T})| + h^2 \Delta t \sum_{k=0}^{N_T-1} \sum_{i,j} m_{i,j}^{k} \left| g_q(x_{i,j}, [\nabla_h u^{k+1}]_{i,j}) \right|^2 \right).
\end{aligned}
\tag{4.16}
$$

Proof.

Step 1. Take $\epsilon > 0$ and consider $\hat{m}_{i,j}^n = m_{i,j}^n + \epsilon$. Note that $\hat{m}_{i,j}^n > 0$ for all i,j,n. Multiply the second equation of (2.11) by $\ln(\hat{m}_{i,j}^n)$ and sum for all i,j and $k = n, \dots, N_T - 1$:

$$
\begin{aligned}
0 =& h^2 \sum_{k=n}^{N_T-1} \sum_{i,j} \left(\hat{m}_{i,j}^{k+1} - \hat{m}_{i,j}^{k}\right) \ln(\hat{m}_{i,j}^{k}) - \frac{\nu h^2 \Delta t}{2} \sum_{k=n}^{N_T-1} \sum_{i,j} \left[\nabla_h \hat{m}^k\right]_{i,j} \cdot \left[\nabla_h \ln(\hat{m}^k)\right]_{i,j}\\
&+ h^2 \Delta t \sum_{k=n}^{N_T-1} \sum_{i,j} \mathcal{T}_{i,j}(u^{k+1}, m^k) \ln(\hat{m}_{i,j}^k).
\end{aligned}
$$

From (4.12) and (4.14), we deduce that

$$
\begin{aligned}
& h^2 \sum_{i,j} \hat{m}_{i,j}^n \ln(\hat{m}_{i,j}^n) + (\nu - \eta) \frac{h^2 \Delta t}{2} \sum_{k=n}^{N_T-1} \sum_{i,j} \left[\nabla_h \hat{m}^k\right]_{i,j} \cdot \left[\nabla_h \ln(\hat{m}^k)\right]_{i,j}\\
\le & 1 + h^2 \sum_{i,j} \hat{m}_{i,j}^{N_T} \ln(\hat{m}_{i,j}^{N_T}) + \frac{h^2 \Delta t}{2\eta} \sum_{k=n}^{N_T-1} \sum_{i,j} \hat{m}_{i,j}^k \left| g_q(x_{i,j}, [\nabla_h u^{k+1}]_{i,j}) \right|^2,
\end{aligned}
$$

and since $\hat{m}_{i,j}^k \ln(\hat{m}_{i,j}^k) \ge -e^{-1}$,

$$
\begin{aligned}
& \max_n h^2 \sum_{i,j} \hat{m}_{i,j}^n |\ln(\hat{m}_{i,j}^n)| + \frac{\nu - \eta}{2} h^2 \Delta t \sum_{k=0}^{N_T-1} \sum_{i,j} \left[\nabla_h \hat{m}^k\right]_{i,j} \cdot \left[\nabla_h \ln(\hat{m}^k)\right]_{i,j}\\
\le \quad & C + h^2 \sum_{i,j} \hat{m}_{i,j}^{N_T} |\ln(\hat{m}_{i,j}^{N_T})| + \frac{h^2 \Delta t}{2\eta} \sum_{k=0}^{N_T-1} \sum_{i,j} \hat{m}_{i,j}^k \left| g_q(x_{i,j}, [\nabla_h u^{k+1}]_{i,j}) \right|^2.
\end{aligned}
\tag{4.17}
$$

Consider now the quantity $|\sqrt{\hat{m}^k}|^2_{H^1(\mathbb{T}_h^2)}$, i.e.

$$
|\sqrt{\hat{m}^k}|^2_{H^1(\mathbb{T}_h^2)} = \sum_{i,j} \left(\sqrt{\hat{m}_{i+1,j}^k} - \sqrt{\hat{m}_{i,j}^k}\right)^2 + \sum_{i,j} \left(\sqrt{\hat{m}_{i,j+1}^k} - \sqrt{\hat{m}_{i,j}^k}\right)^2.
$$

Since $\hat{m}_{i,j}^k > 0$, we can write $\left(\sqrt{\hat{m}_{i+1,j}^k} - \sqrt{\hat{m}_{i,j}^k}\right)^2 = \hat{m}_{i,j}^k \left(\sqrt{1 + h\frac{(D_1^+ \hat{m}^k)_{i,j}}{\hat{m}_{i,j}^k}} - 1\right)^2$ where $(D_1^+ \hat{m}^k)_{i,j}$ is defined in (2.1). Since the inequality $(\sqrt{1+z} - 1)^2 \le z \ln(1+z)$ holds for any number $z \ge -1$, we infer that

$$
\begin{aligned}
\left(\sqrt{\hat{m}_{i+1,j}^k} - \sqrt{\hat{m}_{i,j}^k}\right)^2 &\le h (D_1^+ \hat{m}^k)_{i,j} \ln\left(1 + h \frac{(D_1^+ \hat{m})_{i,j}}{\hat{m}_{i,j}^k}\right)\\
&= h (D_1^+ \hat{m}^k)_{i,j} \left(\ln(\hat{m}_{i+1,j}^k) - \ln(\hat{m}_{i,j}^k)\right)\\
&= h^2 (D_1^+ \hat{m}^k)_{i,j} (D_1^+ \ln(\hat{m}^k))_{i,j}.
\end{aligned}
$$

Since the same kind of estimate holds for $\left(\sqrt{\hat{m}_{i,j+1}^k} - \sqrt{\hat{m}_{i,j}^k}\right)^2$, we obtain that

$$
|\sqrt{\hat{m}^k}|^2_{H^1(\mathbb{T}_h^2)} \le h^2 \sum_{i,j} \left(D_h \hat{m}^k\right)_{i,j} \cdot \left(D_h \ln(\hat{m}^k)\right)_{i,j}, \tag{4.18}
$$

and the fact that $\hat{m}$ satisfies (4.15) follows from (4.17) and (4.18).
Let us now prove (4.16): consider $\alpha \in [1,2)$: there exists a unique number $p \geq 1$ such that $\frac{1}{\alpha} = \frac{1}{2} + \frac{1}{2p}$: we have the interpolation inequality $\|\hat{m}^k\|_{L^\alpha(\mathbb{T}_h^2)} \leq \|\hat{m}^k\|^{\frac{1}{2}}_{L^1(\mathbb{T}_h^2)} \|\hat{m}^k\|^{\frac{1}{2}}_{L^p(\mathbb{T}_h^2)}$. But $\|\hat{m}^k\|_{L^1(\mathbb{T}_h^2)} = 1+\epsilon$ and $\|\hat{m}^k\|^{\frac{1}{2}}_{L^p(\mathbb{T}_h^2)} = \|\sqrt{\hat{m}^k}\|_{L^{2p}(\mathbb{T}_h^2)}$. From the discrete Sobolev inequalities, we deduce that

$$\|\hat{m}^k\|^\alpha_{L^\alpha(\mathbb{T}_h^2)} \leq (1+\epsilon)^{\frac{\alpha}{2}} \|\sqrt{\hat{m}^k}\|^\alpha_{L^{2p}(\mathbb{T}_h^2)} \leq C \|\sqrt{\hat{m}^k}\|^\alpha_{H^1(\mathbb{T}_h^2)} \leq C\left(1 + \|\sqrt{\hat{m}^k}\|^2_{H^1(\mathbb{T}_h^2)}\right),$$

which yields that $\hat{m}$ satisfies (4.16) by summing for all k and using (4.15).

Step 2. We obtain that m satisfies (4.15) and (4.16) by letting ϵ tend to 0. ◻

COROLLARY 4.6. *With the same assumptions as in Lemma 4.5, for any $\alpha \in [1, 4/3)$, there exists a constant c such that*

$$\begin{aligned} &\|D_h m\|^\alpha_{L^\alpha(Q_{h,\Delta t})} + \Delta t \sum_{k=0}^{N_T-1} \left\| \frac{m^{k+1}-m^k}{\Delta t} \right\|^\alpha_{W^{-1,\alpha}(\mathbb{T}_h^2))} \\ \leq & c\left(1 + h^2 \sum_{i,j} m_{i,j}^{N_T} |\ln(m_{i,j}^{N_T})| + h^2 \Delta t \sum_{k=0}^{N_T-1} \sum_{i,j} m_{i,j}^k \left| g_q(x_{i,j}, [\nabla_h u^{k+1}]_{i,j}) \right|^2 \right). \end{aligned} \tag{4.19}$$

Proof. Take $\alpha \in [1,4/3)$. We start by observing that

$$\|D_h m\|^\alpha_{L^\alpha(Q_{h,\Delta t})} \leq C h^2 \Delta t \sum_{k=0}^{N_T-1} \sum_{i,j} |D_1^+ m^k|^\alpha_{i,j} + |D_2^+ m^k|^\alpha_{i,j}.$$

Let us estimate $\sum_{k=0}^{N_T-1} \sum_{i,j} |D_1^+ m^k|^\alpha_{i,j}$ by some quantity depending on $\sum_{k=0}^{N_T-1} \sum_{i,j} (D_1^+ m^k)_{i,j} (D_1^+ \ln(m^k))_{i,j}$: a Hölder inequality yields that

$$\begin{aligned} &\sum_{k=0}^{N_T-1} \sum_{i,j} |D_1^+ m^k|^\alpha_{i,j} \\ \leq & \left(\sum_{k=0}^{N_T-1} \sum_{i,j} |D_1^+ m^k|^2_{i,j} \frac{(D_1^+ \ln(m^k))_{i,j}}{(D_1^+ m^k)_{i,j}} \right)^{\frac{\alpha}{2}} \left(\sum_{k=0}^{N_T-1} \sum_{i,j} \left(\frac{(D_1^+ m^k)_{i,j}}{(D_1^+ \ln(m^k))_{i,j}} \right)^{\frac{\alpha}{2-\alpha}} \right)^{1-\frac{\alpha}{2}} \\ = & \left(\sum_{k=0}^{N_T-1} \sum_{i,j} (D_1^+ m^k)_{i,j} (D_1^+ \ln(m^k))_{i,j} \right)^{\frac{\alpha}{2}} \left(\sum_{k=0}^{N_T-1} \sum_{i,j} \left(\frac{(D_1^+ m^k)_{i,j}}{(D_1^+ \ln(m^k))_{i,j}} \right)^{\frac{\alpha}{2-\alpha}} \right)^{1-\frac{\alpha}{2}}. \end{aligned}$$

Standard calculus yields that $\frac{(D_1^+ m^k)_{i,j}}{(D_1^+ \ln(m^k))_{i,j}} \leq \max(m^k_{i,j}, m^k_{i+1,j}) \leq m^k_{i,j} + m^k_{i+1,j}$, therefore

$$\begin{aligned} &h^2 \Delta t \sum_{k=0}^{N_T-1} \sum_{i,j} |D_1^+ m^k|^\alpha_{i,j} \\ \leq & h^2 \Delta t \left(\sum_{k=0}^{N_T-1} \sum_{i,j} (D_1^+ m^k)_{i,j} (D_1^+ \ln(m^k))_{i,j} \right)^{\frac{\alpha}{2}} \left(\sum_{k=0}^{N_T-1} \sum_{i,j} (m^k_{i,j} + m^k_{i+1,j})^{\frac{\alpha}{2-\alpha}} \right)^{1-\frac{\alpha}{2}} \\ \leq & C \|m\|^{\frac{\alpha}{2}}_{L^{\frac{\alpha}{2-\alpha}}(Q_{h,\Delta t})} \left(h^2 \Delta t \sum_{k=0}^{N_T-1} \sum_{i,j} (D_1^+ m^k)_{i,j} (D_1^+ \ln(m^k))_{i,j} \right)^{\frac{\alpha}{2}}. \end{aligned}$$

Note that $1 \le \frac{\alpha}{2-\alpha} < 2$. The estimate on $\|D_h m\|^\alpha_{L^\alpha(Q_{h,\Delta t})}$ follows by using the same argument for bounding $\sum_{k=0}^{N_T-1}\sum_{i,j}|D_2^+ m^k|^\alpha_{i,j}$ and then (4.17) and (4.16). From (2.9), we deduce that for all grid functions w,

$$
\begin{aligned}
&h^2\sum_{i,j} w_{i,j}\frac{m^{n+1}_{i,j}-m^n_{i,j}}{\Delta t}\\
=&\nu h^2\sum_{i,j}(\nabla_h m^n)_{i,j}\cdot(\nabla_h w)_{i,j} + h^2\sum_{i,j} m^n_{i,j} g_q(x_{i,j},[\nabla_h u^{n+1}]_{i,j})\cdot[\nabla_h w]_{i,j}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
\left(\frac{m^{n+1}-m^n}{\Delta t}, w\right)_{L^2(\mathbb{T}_h^2)} \le & \ \nu |m^n|_{W^{1,\alpha}(\mathbb{T}_h^2)}|w|_{W^{1,\alpha'}(\mathbb{T}_h^2)}\\
&+\left(h^2\sum_{i,j} m^n_{i,j}|g_q(x_{i,j},[\nabla_h u^{n+1}]_{i,j})|^2\right)^{\frac12}\|\sqrt{m^n}\|_{L^{\frac{2\alpha}{2-\alpha}}(\mathbb{T}_h^2)}|w|_{W^{1,\alpha'}(\mathbb{T}_h^2)},
\end{aligned}
$$

and

$$
\begin{aligned}
\left(\Delta t\sum_{k=0}^{N_T-1}\left\|\frac{m^{k+1}-m^k}{\Delta t}\right\|^\alpha_{W^{-1,\alpha}(\mathbb{T}_h^2)}\right)^{\frac1\alpha} \le & \ \nu\|D_h m\|_{L^\alpha(Q_{h,\Delta t})}\\
&+\left(h^2\Delta t\sum_{k=0}^{N_T-1}\sum_{i,j} m^k_{i,j}\left|g_q\left(x_{i,j},\left[\nabla_h u^{k+1}\right]_{i,j}\right)\right|^2\right)^{\frac12}\|m\|^{\frac12}_{L^{\frac{\alpha}{2-\alpha}}(Q_{h,\Delta t})}.
\end{aligned}
\tag{4.20}
$$

Note that $\frac{\alpha}{2-\alpha} < 2$. From (4.16),

$$
\begin{aligned}
&\|m\|^{\frac12}_{L^{\frac{\alpha}{2-\alpha}}(Q_{h,\Delta t})}\\
\le & c\left(1+h^2\sum_{i,j} m^{N_T}_{i,j}|\ln(m^{N_T}_{i,j})| + h^2\Delta t\sum_{k=0}^{N_T-1}\sum_{i,j} m^k_{i,j}\left|g_q(x_{i,j},\left[\nabla_h u^{k+1}\right]_{i,j})\right|^2\right)^{\frac{2-\alpha}{2\alpha}}.
\end{aligned}
\tag{4.21}
$$

The desired estimate on $\Delta t\sum_{k=0}^{N_T-1}\left\|\frac{m^{k+1}-m^k}{\Delta t}\right\|^\alpha_{W^{-1,\alpha}(\mathbb{T}_h^2)}$ follows from (4.20)-(4.21) and from the estimate on $\|D_h m\|^\alpha_{L^\alpha(Q_{h,\Delta t})}$. □

Collecting the above results together with Lemma 4.2, we obtain the following conclusion:

THEOREM 4.7. *If F is continuous and bounded from below by a constant $\underline{F}$, if $(\mathbf{g_1})$, $(\mathbf{g_3})$, $(\mathbf{g_5})$ hold, if u_0 is continuous, then there exists a constant C such that a solution (u,m) of (2.11)- (2.12) satisfies (4.6)-(4.8), and for all $\alpha\in(1,4/3)$,*

$$
\|D_h m\|^\alpha_{L^\alpha(Q_{h,\Delta t})} + \left(\Delta t\sum_{k=0}^{N_T-1}\left\|\frac{m^{k+1}-m^k}{\Delta t}\right\|^\alpha_{W^{-1,\alpha}(\mathbb{T}_h^2))}\right)^{\frac1\alpha} \le C. \tag{4.22}
$$

5. L^1-compactness results. In this section we prove the L^1-compactness of $D_h u$ whenever the discrete heat equation has bounded L^1 data. More precisely, we assume that $u=(u^n)_{n=0,\dots,N_T}$ satisfies

$$
\frac{u^{n+1}_{i,j}-u^n_{i,j}}{\Delta t} - \nu(\Delta_h u^{n+1})_{i,j} = f^n_{i,j} \tag{5.1}
$$

for all $0 \le i,j < N_h$ and all n, $0 \le n < N_T$, where the data $f = (f^n_{i,j})$ and the initial conditions $u^0 = (u^0_{i,j})$ are supposed to satisfy

$$S\|u^0\|_{L^1(\mathbb{T}_h^2)} + \|f\|_{L^1(Q_{h,\Delta t})} \le c \tag{5.2}$$

for some c independent of h and Δt. In what follows, we reconstruct functions on Q from the grid functions u, and we prove the convergence of these functions as h and Δt tend to 0, at least for subsequences. Lemma 5.1 below is concerned with piecewise constant functions built using u. It is similar to results that can be found in Gallouët et al, see e.g. [13, 12, 14] in the context of finite volume methods. Lemma 5.2 deals with approximations of the gradient with respect to x. It seems new to the best of our knowledge and may have an independent interest.

LEMMA 5.1. *Let $u_{h,\Delta t}$ be the piecewise constant function which takes the value $u^{n+1}_{i,j}$ in $(t_n,t_{n+1}) \times (ih-h/2, ih+h/2) \times (jh-h/2, jh+h/2)$. There exists a subsequence of h and Δt (not relabeled) and a function $\tilde u$ such that $u_{h,\Delta t} \to \tilde u$ in in $L^\beta(Q)$ for all $\beta \in [1,2)$. Moreover, $\tilde u \in L^\alpha(0,T;W^{1,\alpha}(\mathbb{T}^2))$ for any $\alpha \in [1,\frac{4}{3})$, and there exist a bounded Radon measure $\tilde\mu$ in Q and a bounded Radon measure $\tilde\mu_0$ in $\mathbb{T}^2$ such that $\tilde u$ is the unique solution of*

$$\begin{cases} \partial_t \tilde u - \nu\Delta\tilde u = \tilde\mu & \text{in } Q, \\ \tilde u(0,\cdot) = \tilde\mu^0 & \text{in } \mathbb{T}^2. \end{cases} \tag{5.3}$$

Proof. Using the L^1 bounds on the data, we may show with the same argument as in § 4.3 that

$$\|u\|_{L^\beta(Q_{h,\Delta t})} + \left(\Delta t \sum_{n=1}^{N_T} |D_h u^n|^\alpha_{L^\alpha(\mathbb{T}_h^2)}\right)^{\frac{1}{\alpha}} \le c$$

for any $\beta \in [1,2)$ and $\alpha \in [1,\frac{4}{3})$. From this estimate and (5.1), we deduce that $\Delta t \sum_{n=1}^{N_T} \|\frac{u^{n+1}-u^n}{\Delta t}\|_{W^{-1,\alpha}(\mathbb{T}_h^2)}$ is uniformly bounded.
Recall that $u_{h,\Delta t}$ is the piecewise constant function which takes the value $u^{n+1}_{i,j}$ in $(t_n,t_{n+1}) \times (ih-h/2, ih+h/2) \times (jh-h/2, jh+h/2)$. We can apply the discrete Aubin-Simon lemma in [14] (Theorem 3.1): up to the extraction of a subsequence, $u_{h,\Delta t}$ converges to a function $\tilde u$ in $L^1(Q)$, and in fact in $L^\beta(Q)$ for all $\beta \in [1,2)$. Moreover, $\tilde u \in L^\alpha(0,T;W^{1,\alpha}(\mathbb{T}^2))$ for any $\alpha \in [1,\frac{4}{3})$.
Let $f_{h,\Delta t}$ be the piecewise constant function on which takes the values $f^n_{i,j}$ in $(t_n,t_{n+1}) \times (ih-h/2, ih+h/2) \times (jh-h/2, jh+h/2)$. Up to the extraction of a subsequence, $f_{h\Delta t}$ converges in the weak-$*$ topology to some bounded Radon measure $\tilde\mu$ on Q. Call u^0_h the piecewise constant function on $\mathbb{T}^2$ which takes the values $u^0_{i,j}$ in $(ih-h/2, ih+h/2) \times (jh-h/2, jh+h/2)$. We may assume that u^0_h converges to a bounded measure $\tilde\mu^0$ on $\mathbb{T}^2$. In particular, testing (5.1) with smooth functions and passing to the limit, this implies that $\tilde u$ satisfies

$$\int_0^T \int_{\mathbb{T}^2} \tilde u \left(-\varphi_t - \nu\Delta\varphi\right) dxdt = \int_0^T \int_{\mathbb{T}^2} \varphi \, d\tilde\mu + \int_{\mathbb{T}^2} \varphi(0) d\tilde\mu^0 \,,$$

for every $\varphi \in C^2(\overline{Q})$ such that $\varphi(T) = 0$. Notice that $\tilde u$ is the unique solution of the above weak formulation. ☐

We now define an approximation of $D\tilde u$ from the grid function u. For a real number z, let floor(z) be the largest integer that does not exceed z, ceil(z) be the smallest integer that is not less than z. Let $\widetilde{Du}_{h,\Delta t}$ be the piecewise constant function from Q to $\mathbb{R}^2$ which takes the value

$$\left(\left(D_1^+ u^{n+1}\right)_{\mathrm{floor}(\frac{i}{2}),\mathrm{ceil}(\frac{j}{2})}, \left(D_2^+ u^{n+1}\right)_{\mathrm{ceil}(\frac{i}{2}),\mathrm{floor}(\frac{j}{2})} \right)$$

in $(t_n, t_{n+1}) \times \left(i\frac{h}{2}, (i+1)\frac{h}{2}\right) \times \left(j\frac{h}{2}, (j+1)\frac{h}{2}\right)$. More explicitly, $\widetilde{Du}_{h,\Delta t}$ takes the value

$$
\begin{array}{lll}
\left(\left(D_1^+ u^{n+1}\right)_{i,j}, \left(D_2^+ u^{n+1}\right)_{i,j}\right) & \text{in} & (t_n, t_{n+1}) \times \left(ih, ih + \frac{h}{2}\right) \times \left(jh, jh + \frac{h}{2}\right) \\
\left(\left(D_1^+ u^{n+1}\right)_{i,j}, \left(D_2^+ u^{n+1}\right)_{i,j-1}\right) & \text{in} & (t_n, t_{n+1}) \times \left(ih, ih + \frac{h}{2}\right) \times \left(jh - \frac{h}{2}, jh\right) \\
\left(\left(D_1^+ u^{n+1}\right)_{i-1,j}, \left(D_2^+ u^{n+1}\right)_{i,j-1}\right) & \text{in} & (t_n, t_{n+1}) \times \left(ih - \frac{h}{2}, ih\right) \times \left(jh - \frac{h}{2}, jh\right) \\
\left(\left(D_1^+ u^{n+1}\right)_{i-1,j}, \left(D_2^+ u^{n+1}\right)_{i,j}\right) & \text{in} & (t_n, t_{n+1}) \times \left(ih - \frac{h}{2}, ih\right) \times \left(jh, jh + \frac{h}{2}\right).
\end{array}
$$

LEMMA 5.2. *Up to the extraction of a subsequence, the functions $\widetilde{Du}_{h,\Delta t}$ converge a.e. to $D\tilde{u}$ in Q, and in $L^\alpha(Q)$ for any $\alpha \in [1, \frac{4}{3})$.*

Proof. Since $\tilde{u}$ is the unique weak solution of (5.3), for every sequence of smooth functions $\tilde{\mu}_\delta$ and smooth initial data $\tilde{\mu}_\delta^0$ converging to $\tilde{\mu}$ and to $\tilde{\mu}^0$ respectively, in the weak-$*$ sense of measures, (which, for instance, can be constructed by convolution), the smooth solutions $\tilde{U}_\delta$ satisfying

$$
\begin{cases}
\partial_t \tilde{U}_\delta - \nu \Delta \tilde{U}_\delta = \tilde{\mu}_\delta & \text{in } Q, \\
\tilde{U}_\delta(0, \cdot) = \tilde{\mu}_\delta^0 & \text{in } \mathbb{T}^2
\end{cases}
$$

will converge to $\tilde{u}$, e.g. in $L^\alpha(0, T; W^{1,\alpha}(\mathbb{T}^2))$ for any $\alpha \in [1, \frac{4}{3})$, see e.g. [7].
We now consider the finite difference approximation

$$
\frac{U_{\delta,i,j}^{n+1} - U_{\delta,i,j}^n}{\Delta t} - \nu(\Delta_h U_\delta^{n+1})_{i,j} = \mu_{\delta,i,j}^n, \tag{5.4}
$$

with $(U_\delta^0)_{i,j} = \tilde{\mu}_\delta^0(x_{i,j})$ and $\mu_{\delta,i,j}^n = \tilde{\mu}_\delta(t_n, x_{i,j})$.
Let σ be a positive real number: let T_σ be the piecewise linear function defined on $\mathbb{R}$ by $T_\sigma(z) = \max(-\sigma, \min(z, \sigma))$. Let the grid function e on $Q_{h,\Delta t}$ be given by $e_{i,j}^n = u_{i,j}^n - U_{\delta,i,j}^n$. Define also, for $0 \le i, j < N_h$,

$$
\mathcal{I}_j^n = \left\{ i \text{ s.t. } \max[|e_{i+1,j}^n|, |e_{i,j}^n|] \le \sigma \right\}; \qquad \mathcal{J}_i^n = \left\{ j \text{ s.t. } \max[|e_{i,j+1}^n|, |e_{i,j}^n|] \le \sigma \right\}.
$$

For any $s : 0 < s < 1$ we have

$$
h^2 \Delta t \sum_{n=0}^{N_T - 1} \sum_{i,j} |D_h e_{i,j}^{n+1}|^s \le h^2 \Delta t \sum_{n=0}^{N_T - 1} \sum_{i,j} |D_1^+ e_{i,j}^{n+1}|^s + |D_2^+ e_{i,j}^{n+1}|^s.
$$

The first term can be estimated as follows

$$
\begin{aligned}
h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j}|D_1^+e_{i,j}^{n+1}|^s \le{}& h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{j}\sum_{i\in\mathcal{I}_j^{n+1}}(D_1^+e_{i,j}^{n+1}D_1^+T_\sigma(e^{n+1})_{i,j})^{\frac{s}{2}}\\
&+h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{j}\sum_{i\notin\mathcal{I}_j^{n+1}}|D_1^+e_{i,j}^{n+1}|^s\\
\le{}& T^{1-\frac{s}{2}}\left(h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j}D_1^+e_{i,j}^{n+1}D_1^+T_\sigma(e^{n+1})_{i,j}\right)^{\frac{s}{2}}\\
&+\left(h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j}|D_1^+e_{i,j}^{n+1}|\right)^s\left(h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{j}\sum_{i\notin\mathcal{I}_j^{n+1}}1\right)^{1-s}\\
\le{}& T^{1-\frac{s}{2}}\left(h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j}D_1^+e_{i,j}^{n+1}D_1^+T_\sigma(e^{n+1})_{i,j}\right)^{\frac{s}{2}}\\
&+\left(h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j}|D_1^+e_{i,j}^{n+1}|\right)^s\left(h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j}\frac{|e_{i+1,j}^{n+1}|+|e_{i,j}^{n+1}|}{\sigma}\right)^{1-s}.
\end{aligned}
$$

Similarly we estimate the term with D_2^+ using the set $\mathcal{J}_i^n$, and overall we deduce that

$$
\begin{aligned}
h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j}|D_he_{i,j}^{n+1}|^s \le{}& c\,T^{1-\frac{s}{2}}\left(h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j}D_he_{i,j}^{n+1}\cdot D_hT_\sigma(e^{n+1})_{i,j}\right)^{\frac{s}{2}}\\
&+c\|D_he\|_{L^1(Q_{h,\Delta t})}^s\|e\|_{L^1(Q_{h,\Delta t})}^{1-s}\sigma^{-(1-s)}
\end{aligned}
\tag{5.5}
$$

for some constant c only depending on s. We estimate the first term from the discrete equation

$$
\begin{aligned}
\nu\,h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j}D_he_{i,j}^{n+1}\cdot D_hT_\sigma(e^{n+1})_{i,j} ={}& h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j}(f_{i,j}^n-\mu_{\delta,i,j}^n)T_\sigma(e_{i,j}^{n+1})\\
&-h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j}\frac{e_{i,j}^{n+1}-e_{i,j}^n}{\Delta t}T_\sigma(e_{i,j}^{n+1})
\end{aligned}
$$

which implies, using that $(x-y)T_\sigma(x)\ge\Theta_\sigma(x)-\Theta_\sigma(y)$ for the nonnegative and convex function $\Theta(s)=\int_0^s T_\sigma(r)dr$,

$$
\nu\,h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j}D_he_{i,j}^{n+1}\cdot D_hT_\sigma(e^{n+1})_{i,j}\le\sigma\left(\|f-\mu_\delta\|_{L^1(Q_{h,\Delta t})}+\|e^0\|_{L^1(\mathbb{T}_h^2)}\right).
$$

Therefore, we deduce from (5.5)

$$
\begin{aligned}
h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j}|D_he_{i,j}^{n+1}|^s\le{}& c\,T^{1-\frac{s}{2}}\sigma^{\frac{s}{2}}\left(\|f-\mu_\delta\|_{L^1(Q_{h,\Delta t})}+\|e^0\|_{L^1(\mathbb{T}_h^2)}\right)^{\frac{s}{2}}\\
&+c\|D_he\|_{L^1(Q_{h,\Delta t})}^s\|e\|_{L^1(Q_{h,\Delta t})}^{1-s}\sigma^{-(1-s)}.
\end{aligned}
$$

Taking the minimum of the right hand side w.r.t. σ, and using the L^1 bounds for μ_δ, $\mu_{0\delta}$ and the data in (5.2), we see that

$$\||D_h e|^s\|_{L^1(Q_{h,\Delta t})} \le c\|D_h e\|^{\theta s}_{L^1(Q_{h,\Delta t})}\|e\|^{\theta(1-s)}_{L^1(Q_{h,\Delta t})}$$

for some c and θ depending on s but not on h or δ. Recalling the definition of e, and the estimate on the discrete gradient, we have proved that

$$\||D_h u - D_h U_\delta|^s\|_{L^1(Q_{h,\Delta t})} \le c\|u - U_\delta\|^{\theta(1-s)}_{L^1(Q_{h,\Delta t})}.$$

Hence,

$$\|\,|D_h u - D_h U_\delta|^s\|_{L^1(Q_{h,\Delta t})} \le c\|u_{h,\Delta t} - U_{\delta,h,\Delta t}\|^{\theta(1-s)}_{L^1(Q_{h,\Delta t})}, \tag{5.6}$$

where $u_{h,\Delta t}$ has been defined in Lemma 5.1 and $U_{\delta,h,\Delta t}$ is the piecewise constant function that takes the value $U^{n+1}_{\delta,i,j}$ in $(t_n,t_{n+1})\times(ih-h/2,ih+h/2)\times(jh-h/2,jh+h/2)$.

Let us also define $\widetilde{DU}_{\delta,h,\Delta t}$ from the grid function U_δ in a similar way as $\widetilde{Du}_{h,\Delta t}$: it takes the values

$$\left(\left(D_1^+U_\delta^{n+1}\right)_{\mathrm{floor}(\frac{i}{2}),\mathrm{ceil}(\frac{j}{2})},\left(D_2^+U_\delta^{n+1}\right)_{\mathrm{ceil}(\frac{i}{2}),\mathrm{floor}(\frac{j}{2})}\right)$$

in $(t_n,t_{n+1})\times\left(i\frac{h}{2},(i+1)\frac{h}{2}\right)\times\left(j\frac{h}{2},(j+1)\frac{h}{2}\right)$. Therefore, we see that

$$\begin{aligned}
&\|\,|\widetilde{Du}_{h,\Delta t} - D\tilde u|^s\|_{L^1(Q)}\\
\le&\|\,|\widetilde{Du}_{h,\Delta t} - \widetilde{DU}_{\delta,h,\Delta t}|^s\|_{L^1(Q)} + \|\,|\widetilde{DU}_{\delta,h,\Delta t} - D\tilde U_\delta|^s\|_{L^1(Q)} + \|\,|D\tilde U_\delta - D\tilde u|^s\|_{L^1(Q)}\\
\le&c\|u_{h,\Delta t} - U_{\delta,h,\Delta t}\|^{\theta(1-s)}_{L^1(Q_{h,\Delta t})} + \|\,|\widetilde{DU}_{\delta,h,\Delta t} - D\tilde U_\delta|^s\|_{L^1(Q)} + \|\,|D\tilde U_\delta - D\tilde u|^s\|_{L^1(Q)}\\
\le&c\left(\|u_{h,\Delta t} - \tilde u\|^{\theta(1-s)}_{L^1(Q_{h,\Delta t})} + \|\tilde u - \tilde U_\delta\|^{\theta(1-s)}_{L^1(Q_{h,\Delta t})} + \|\tilde U_\delta - U_{\delta,h,\Delta t}\|^{\theta(1-s)}_{L^1(Q_{h,\Delta t})}\right)\\
&+\|\,|\widetilde{DU}_{\delta,h,\Delta t} - D\tilde U_\delta|^s\|_{L^1(Q)} + \|\,|D\tilde U_\delta - D\tilde u|^s\|_{L^1(Q)}
\end{aligned}$$

where we have used (5.6) to obtain the third line. At fixed δ, since $\tilde U_\delta$ is a smooth solution of the heat equation, the discrete approximation $U_{\delta,h,\Delta t}$ converges to $\tilde U_\delta$ in $L^2(Q)$ and $\widetilde{DU}_{\delta,h,\Delta t}$ converges to $D\tilde U_\delta$ in $L^2(Q;\mathbb{R}^2)$. Using also Lemma 5.1, we get that

$$\limsup_{h,\Delta t\to 0}\||Du_{h,\Delta t} - D\tilde u|^s\|_{L^1(Q)} \le c\|\tilde u - \tilde U_\delta\|^{\theta(1-s)}_{L^1(Q)} + \||D\tilde U_\delta - D\tilde u|^s\|_{L^1(Q)}.$$

We conclude using the strong convergence of $\tilde U_\delta$ to $\tilde u$ in $L^\alpha(0,T;W^{1,\alpha}(\mathbb{T}^2))$ for any $\alpha\in[1,\frac{4}{3})$ (see e.g. [7], as $\delta\to 0$. So

$$\|\,|\widetilde{Du}_{h,\Delta t} - D\tilde u|^s\|_{L^1(Q)} \overset{\Delta t,h\to 0}{\to} 0$$

which in particular implies that $\widetilde{Du}_{h,\Delta t}$ converges to $D\tilde u$ a.e. in Q and then, by Vitali's theorem, in $L^\alpha(Q)$ for any $\alpha\in[1,\frac{4}{3})$. ◻

Remark 4. *As a consequence of Lemma 5.2, for any $\xi\in\mathbb{R}^2$, there exists a subsequence of h and Δt (not relabeled) such that the maps $(t,x)\mapsto\widetilde{Du}_{h,\Delta t}(t,x+h\xi)$ also converge to $D\tilde u$ a.e. and in $L^\alpha(Q)$ for any $\alpha\in[1,\frac{4}{3})$.*

Remark 5. *Alternative strategies can be used to construct a function defined on Q from the grid function u. For example, we can define $w_{h,\Delta t}$ as the continuous and piecewise trilinear function on $\bar Q$ which takes the values $u^{\max(1,n)}_{i,j}$ at $(t_n,x_{i,j})$ and which is trilinear in the rectangles of the time-space grid $Q_{h,\Delta t}$. The* *advantage of taking $w_{h,\Delta t}$ instead of $u_{h,\Delta t}$ is that the former has weakly integrable partial derivatives with respect to the spatial variable. Therefore, we can use directly $Dw_{h,\Delta t}$ instead of having to define an independent approximation of $D\tilde{u}$ such as $\widetilde{Du}_{h,\Delta t}$. It is then possible to prove the following lemma, which may replace both Lemmas 5.1 and 5.2:*

LEMMA 5.3. *There exists a subsequence of h and Δt (not relabeled) and a function $\tilde{u}$ such that $w_{h,\Delta t} \to \tilde{u}$ in $L^\alpha(0,T;W^{1,\alpha}(\mathbb{T}^2))$ for any $\alpha \in [1,\frac{4}{3})$ and in $L^\beta(Q)$ for all $\beta \in [1,2)$. In particular, $w_{h,\Delta t} \to \tilde{u}$ and $Dw_{h,\Delta t} \to D\tilde{u}$ in $L^1(Q)$ and almost everywhere in Q.*

Proof. *The strategy of proof is similar except that we may directly use the continuous version of the compactness lemma of Aubin-Simon, see [29], for the function $w_{h,\Delta t}$.* ◻

6. From the discrete to the continuous system.

6.1. A priori estimates and compactness. Let $u_{h,\Delta t}$ and $\widetilde{Du}_{h,\Delta t}$ be the piecewise constant functions defined in Lemmas 5.1 and 5.2 respectively: up to the extraction of a subsequence, we can assume that $u_{h,\Delta t} \to \tilde{u}$ in $L^\beta(Q)$ for all $\beta \in [1,2)$ and that $\widetilde{Du}_{h,\Delta t}$ a.e. to $D\tilde{u}$ in Q and in $L^\alpha(Q)$ for any $\alpha \in [1,\frac{4}{3})$.
Let $m_{h,\Delta t}$ be the piecewise constant function which takes the value $m^n_{i,j}$ in $(t_n,t_{n+1})\times(ih-h/2,ih+h/2)\times(jh-h/2,jh+h/2)$, and $\widetilde{Dm}_{h,\Delta t}$ be the piecewise constant function from Q to $\mathbb{R}^2$ which takes the value

$$\left(\left(D_1^+ m^n\right)_{\mathrm{floor}(\frac{i}{2}),\mathrm{ceil}(\frac{j}{2})}, \left(D_2^+ m^n\right)_{\mathrm{ceil}(\frac{i}{2}),\mathrm{floor}(\frac{j}{2})} \right)$$

in $(t_n,t_{n+1}) \times \left(i\frac{h}{2},(i+1)\frac{h}{2}\right) \times \left(j\frac{h}{2},(j+1)\frac{h}{2}\right)$. From Theorem 4.7, we may also assume that $m_{h,\Delta t} \to \tilde{m}$ in $L^1(Q)$ and almost everywhere in Q, and that $\widetilde{Dm}_{h,\Delta t} \to D\tilde{m}$ weakly in $L^\alpha(Q)$ for any $\alpha \in [1,\frac{4}{3})$. Moreover, for all $\eta>0$, there exists a constant c_η such that for all $z \ge 0$, $F(z) \le \frac{zF(z)}{\eta} + c_\eta$. This fact and estimate (4.8) yield the equi-integrability of $F(m_{h,\Delta t})$. By Vitali's theorem, $F(m_{h,\Delta t}) \to F(\tilde{m})$ in $L^1(Q)$.
From the observations above, the piecewise constant function which takes the value

$$\frac{u^{n+1}_{i,j} - u^n_{i,j}}{\Delta t} - \nu(\Delta_h u^{n+1})_{i,j} - F(m^n_{i,j})$$

in $(t_n,t_{n+1})\times(ih-h/2,ih+h/2)\times(jh-h/2,jh+h/2)$ converges to $\frac{\partial \tilde{u}}{\partial t} - \nu\Delta\tilde{u} - F(\tilde{m})$ in the sense of distributions.

6.2. Stability of the discrete Bellman equation. We now pass to the limit in the discrete Bellman equation.

The main difficulty is to handle the nonlinear term $g(x_{i,j},\left[\nabla_h u^{n+1}\right]_{i,j})$; here we wish to use the a.e. convergence of the gradients obtained in § 5. We adapt the method used for continuous problems in [8]. Note that $\left[\nabla_h u^{n+1}\right]_{i,j}$ is the value taken by the piecewise constant function with values in $\mathbb{R}^4$

$$\left(\vec{e_1}\cdot\widetilde{Du}_{h,\Delta t}(\cdot+\tfrac{h}{2}\vec{e_1}), \vec{e_1}\cdot\widetilde{Du}_{h,\Delta t}(\cdot-\tfrac{h}{2}\vec{e_1}), \vec{e_2}\cdot\widetilde{Du}_{h,\Delta t}(\cdot+\tfrac{h}{2}\vec{e_2}), \vec{e_2}\cdot\widetilde{Du}_{h,\Delta t}(\cdot-\tfrac{h}{2}\vec{e_2}) \right)$$

at (t,x) such that $|x_1-ih|<h/2$, $|x_2-jh|<h/2$, $t_n \le t < t_{n+1}$. From the continuity of g, the consistency assumption and Remark 4,

$$g_{h,\Delta t} \to g(x, D_1\tilde{u}, D_1\tilde{u}, D_2\tilde{u}, D_2\tilde{u}) = H(x,D\tilde{u}) \qquad \text{a.e. in } Q,$$

where $g_{h,\Delta t}$ is the piecewise constant function which take the value $g(x_{i,j},\left[\nabla_h u^{n+1}\right]_{i,j})$ for (t,x) such that $|x_1-ih|<h/2$, $|x_2-jh|<h/2$, $t_n \le t < t_{n+1}$.

Let now φ be a smooth function on $\mathbb{T}^2$ such that $\varphi \geq 0$, with $\varphi(T) = 0$. We multiply the discrete Bellman equation by $\varphi(t_{n+1}, x_{i,j})$ and sum for all i, j and $n = 0, \ldots, N_T - 1$. Since, by convexity,

$$g(x, q) \geq g(x, 0) + g_q(x, 0) \cdot q$$

the regularity of g w.r.t. x and the L^1-compactness of $\widetilde{Du}_{h,\Delta t}$ allow us to apply Fatou's lemma obtaining

$$\liminf_{h \to 0} h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} g(x_{i,j}, \left[\nabla_h u^{n+1}\right]_{i,j}) \varphi(t_{n+1}, x_{i,j}) \geq \int_Q H(x, D\tilde{u}) \varphi \, dxdt \,.$$

Passing to the limit in the other terms of the equation, we deduce that

$$-\int_Q \tilde{u}\, \varphi_t \, dxdt + \nu \int_Q D\tilde{u} D\varphi \, dxdt + \int_Q H(x, D\tilde{u}) \varphi \, dxdt \leq \int_{\mathbb{T}^2} u_0 \, \varphi(0) \, dx + \int_Q F(\tilde{m}) \varphi \, dxdt \,.$$

We now wish to obtain the reverse inequality, which is the difficult part. We start by noticing that, since the monotonicity assumption implies

$$g(x, q_1, q_2, q_3, q_4) \leq g(x, -q_1^-, q_2^+, -q_3^-, q_4^+)$$

from (3.1) and (3.2) and the fact that $g(x, 0)$ is bounded, we know there exists $\lambda > 0$ such that

$$g(x, q_1, q_2, q_3, q_4) \leq \nu \, \lambda \left[1 + (q_1^-)^2 + (q_2^+)^2 + (q_3^-)^2 + (q_4^+)^2\right] \,. \tag{6.1}$$

We multiply the discrete Bellman equation by $e^{-\lambda u_{i,j}^{n+1}} \varphi(t_{n+1}, x_{i,j})$ and sum for all i, j and $n = 0, \ldots, N_T - 1$. We obtain

$$\begin{aligned}
&h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} \frac{u_{i,j}^{n+1} - u_{i,j}^n}{\Delta t} \, e^{-\lambda u_{i,j}^{n+1}} \varphi(t_{n+1}, x_{i,j}) \\
&\qquad + \nu \, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} D_h u_{i,j}^{n+1} \cdot D_h (e^{-\lambda u_{i,j}^{n+1}} \varphi(t_{n+1}, x_{i,j}))_{i,j} \\
&\qquad + h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} g(x_{i,j}, \left[\nabla_h u^{n+1}\right]_{i,j}) e^{-\lambda u_{i,j}^{n+1}} \varphi(t_{n+1}, x_{i,j}) \\
&= h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} F(m_{i,j}^n) e^{-\lambda u_{i,j}^{n+1}} \varphi(t_{n+1}, x_{i,j}).
\end{aligned} \tag{6.2}$$

Since u is uniformly bounded below, the last term converges by dominated convergence, so

$$\lim_{h \to 0} h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} F(m_{i,j}^n) e^{-\lambda u_{i,j}^{n+1}} \varphi(t_{n+1}, x_{i,j}) = \int_Q F(\tilde{m}) e^{-\lambda \tilde{u}} \, \varphi \, dxdt. \tag{6.3}$$

By convexity of $s \mapsto e^{-\lambda s}$ and since $\phi(T, \cdot) = 0$, we have

$$\begin{aligned}
&h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} \frac{u_{i,j}^{n+1} - u_{i,j}^n}{\Delta t} \, e^{-\lambda u_{i,j}^{n+1}} \varphi(t_{n+1}, x_{i,j}) \\
&\quad \leq \frac{1}{\lambda} \, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} \frac{e^{-\lambda u_{i,j}^n} - e^{-\lambda u_{i,j}^{n+1}}}{\Delta t} \, \varphi(t_{n+1}, x_{i,j}) \\
&\quad = \frac{1}{\lambda} \, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} e^{-\lambda u_{i,j}^n} \, \frac{\varphi(t_{n+1}, x_{i,j}) - \varphi(t_n, x_{i,j})}{\Delta t} + \frac{1}{\lambda} \, h^2 \sum_{i,j} e^{-\lambda u_{i,j}^0} \varphi(0, x_{i,j}) \,,
\end{aligned}$$

and so, again by dominated convergence,

$$
\begin{aligned}
\limsup_{h\to 0} \quad & h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} \frac{u_{i,j}^{n+1} - u_{i,j}^{n}}{\Delta t} \, e^{-\lambda u_{i,j}^{n+1}} \varphi(t_{n+1}, x_{i,j}) \\
& \le \frac{1}{\lambda} \int_Q e^{-\lambda u} \varphi_t \, dxdt + \frac{1}{\lambda} \int_{\mathbb{T}^2} e^{-\lambda u_0} \, \varphi(0) \, dx \, .
\end{aligned}
\tag{6.4}
$$

Let us deal now jointly with the second and third term in (6.2). First we split the energy term according to the sign of $D_1^+(u)_{i,j}$ (and $D_2^+(u)_{i,j}$, respectively); indeed, we can write

$$
\begin{aligned}
& D_1^+(u^{n+1})_{i,j} D_1^+(e^{-\lambda u^{n+1}} \varphi^{n+1})_{i,j} \\
& = (D_1^+(u^{n+1})_{i,j})^+ \left(e^{-\lambda u_{i+1,j}^{n+1}} - e^{-\lambda u_{i,j}^{n+1}} \right) \varphi_{i+1,j}^{n+1} - (D_1^+(u^{n+1})_{i,j})^- \left(e^{-\lambda u_{i+1,j}^{n+1}} - e^{-\lambda u_{i,j}^{n+1}} \right) \varphi_{i,j}^{n+1} \\
& + (D_1^+(u^{n+1})_{i,j})^+ (\varphi_{i+1,j}^{n+1} - \varphi_{i,j}^{n+1}) e^{-\lambda u_{i,j}^{n+1}} - (D_1^+(u^{n+1})_{i,j})^- (\varphi_{i+1,j}^{n+1} - \varphi_{i,j}^{n+1}) e^{-\lambda u_{i+1,j}^{n+1}}
\end{aligned}
$$

and the same for the term with D_2^+. Reordering the indexes in the sum, this means that the D_1 part in the second order term can be read as

$$
\begin{aligned}
& \nu \, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} (D_1^+(u^{n+1})_{i-1,j})^+ \frac{1}{h} \left(e^{-\lambda u_{i,j}^{n+1}} - e^{-\lambda u_{i-1,j}^{n+1}} \right) \varphi_{i,j}^{n+1} \\
& - \nu \, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} (D_1^+(u^{n+1})_{i,j})^- \frac{1}{h} \left(e^{-\lambda u_{i+1,j}^{n+1}} - e^{-\lambda u_{i,j}^{n+1}} \right) \varphi_{i,j}^{n+1} \\
& + \nu \, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} (D_1^+(u^{n+1})_{i,j})^+ D_1^+(\varphi^{n+1})_{i,j} e^{-\lambda u_{i,j}^{n+1}} \\
& - \nu \, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} (D_1^+(u^{n+1})_{i,j})^- D_1^+(\varphi^{n+1})_{i,j} e^{-\lambda u_{i+1,j}^{n+1}} \, ,
\end{aligned}
$$

which is equal to

$$
\begin{aligned}
& \nu \, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} |(D_1^+(u^{n+1})_{i-1,j})^+|^2 \frac{e^{-\lambda u_{i,j}^{n+1}} - e^{-\lambda u_{i-1,j}^{n+1}}}{u_{i,j}^{n+1} - u_{i-1,j}^{n+1}} \, \varphi_{i,j}^{n+1} \\
& + \nu \, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} |D_1^+(u^{n+1})_{i,j})^-|^2 \frac{e^{-\lambda u_{i+1,j}^{n+1}} - e^{-\lambda u_{i,j}^{n+1}}}{u_{i+1,j}^{n+1} - u_{i,j}^{n+1}} \, \varphi_{i,j}^{n+1} \\
& + \nu \, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} \left\{ (D_1^+(u^{n+1})_{i,j})^+ e^{-\lambda u_{i,j}^{n+1}} - (D_1^+(u^{n+1})_{i,j})^- e^{-\lambda u_{i+1,j}^{n+1}} \right\} D_1^+(\varphi^{n+1})_{i,j} \, .
\end{aligned}
$$

We proceed similarly for the part with D_2. Therefore,

$$
\begin{aligned}
&\nu\, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} D_h u_{i,j}^{n+1} \cdot D_h (e^{-\lambda u_{i,j}^{n+1}} \varphi(t_{n+1}, x_{i,j}))_{i,j} \\
&\qquad + h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} g(x_{i,j}, \left[\nabla_h u^{n+1}\right]_{i,j}) e^{-\lambda u_{i,j}^{n+1}} \varphi(t_{n+1}, x_{i,j}) \\
&= \nu\, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} \left\{ (D_1^+ (u^{n+1})_{i,j})^+ e^{-\lambda u_{i,j}^{n+1}} - (D_1^+ (u^{n+1})_{i,j})^- e^{-\lambda u_{i+1,j}^{n+1}} \right\} D_1^+ (\varphi^{n+1})_{i,j} \\
&+ \nu\, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} \left\{ (D_2^+ (u^{n+1})_{i,j})^+ e^{-\lambda u_{i,j}^{n+1}} - (D_2^+ (u^{n+1})_{i,j})^- e^{-\lambda u_{i,j+1}^{n+1}} \right\} D_2^+ (\varphi^{n+1})_{i,j} \\
&+ \nu\, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} |(D_1^+ (u^{n+1})_{i-1,j})^+|^2 \frac{e^{-\lambda u_{i,j}^{n+1}} - e^{-\lambda u_{i-1,j}^{n+1}}}{u_{i,j}^{n+1} - u_{i-1,j}^{n+1}} \varphi_{i,j}^{n+1} \\
&+ \nu\, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} |D_1^+ (u^{n+1})_{i,j})^-|^2 \frac{e^{-\lambda u_{i+1,j}^{n+1}} - e^{-\lambda u_{i,j}^{n+1}}}{u_{i+1,j}^{n+1} - u_{i,j}^{n+1}} \varphi_{i,j}^{n+1} \\
&+ \nu\, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} |(D_2^+ (u^{n+1})_{i,j-1})^+|^2 \frac{e^{-\lambda u_{i,j}^{n+1}} - e^{-\lambda u_{i,j-1}^{n+1}}}{u_{i,j}^{n+1} - u_{i,j-1}^{n+1}} \varphi_{i,j}^{n+1} \\
&+ \nu\, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} |(D_2^+ (u^{n+1})_{i,j})^-|^2 \frac{e^{-\lambda u_{i,j+1}^{n+1}} - e^{-\lambda u_{i,j}^{n+1}}}{u_{i,j+1}^{n+1} - u_{i,j}^{n+1}} \varphi_{i,j}^{n+1} \\
&\qquad + h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} g(x_{i,j}, \left[\nabla_h u^{n+1}\right]_{i,j}) e^{-\lambda u_{i,j}^{n+1}} \varphi_{i,j}^{n+1}.
\end{aligned}
$$

The first two terms in the right-hand side converge to $\nu \int_0^T \int_{\mathbb{T}_2} e^{-\lambda \tilde{u}} D\tilde{u} \cdot d\phi$ by Lebesgue theorem, since $\widetilde{Du}_{h,\Delta t}$ converges strongly in L^1, φ is smooth and $e^{-\lambda u_h}$ is uniformly bounded and a.e. convergent. As far as the remaining terms are concerned, using that

$$
\frac{e^{-\lambda s} - e^{-\lambda s'}}{s - s'} \leq -\lambda e^{-\lambda \max(s,s')},
$$

and due to (6.1), we observe that the last five terms under summation are bounded above, so that we can again apply Fatou's lemma, on account of the a.e. convergence of $u_{h,\Delta t}$ and $\widetilde{Du}_{h,\Delta t}$. Therefore, we conclude that

$$
\begin{aligned}
\limsup_{h\to 0} \quad & \nu\, h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} D_h u_{i,j}^{n+1} \cdot D_h (e^{-\lambda u_{i,j}^{n+1}} \varphi(t_{n+1}, x_{i,j}))_{i,j} \\
&+ h^2 \Delta t \sum_{n=0}^{N_T-1} \sum_{i,j} g(x_{i,j}, \left[\nabla_h u^{n+1}\right]_{i,j}) e^{-\lambda u_{i,j}^{n+1}} \varphi(t_{n+1}, x_{i,j}) \\
\leq \nu \int_Q & D\tilde{u} D\varphi\, e^{-\lambda \tilde{u}}\, dxdt - \nu\, \lambda \int_Q |D\tilde{u}|^2\, e^{-\lambda \tilde{u}}\, \varphi\, dxdt + H(x, D\tilde{u}) e^{-\lambda \tilde{u}}\, \varphi\, dxdt\,.
\end{aligned}
\tag{6.5}
$$

Putting together (6.3)-(6.4)-(6.5), we deduce from (6.2) that $\tilde{u}$ satisfies

$$\frac{1}{\lambda}\int_Q e^{-\lambda\tilde{u}}\varphi_t\,dxdt + \frac{1}{\lambda}\int_{\mathbb{T}^2} e^{-\lambda u_0}\,\varphi(0)\,dx + \nu\int_Q D\tilde{u}D\varphi\,e^{-\lambda\tilde{u}}\,dxdt \\ -\nu\,\lambda\int_Q |D\tilde{u}|^2\,e^{-\lambda\tilde{u}}\,\varphi\,dxdt + \int_Q H(x,D\tilde{u})e^{-\lambda\tilde{u}}\,\varphi\,dxdt \geq \int_Q F(\tilde{m})e^{-\lambda\tilde{u}}\,\varphi\,dxdt \tag{6.6}$$

for every smooth $\varphi \geq 0$. In order to conclude, we need now to get rid of the exponential in the above inequality (6.6). To this purpose, we first observe that

$$e^{-\lambda\tilde{u}} \in L^2(0,T;H^1(\mathbb{T}^2)) \cap L^\infty(Q). \tag{6.7}$$

This can be easily proved obtaining an a priori estimate on $e^{-\lambda u_{h,\Delta t}}$. Indeed, whenever u is a grid function which solves (5.1) for some data satisfying (5.2), we have

$$\left| h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j}\frac{u_{i,j}^{n+1}-u_{i,j}^n}{\Delta t}\,\psi(u_{i,j}^{n+1}) + \nu\,h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j} D_h u_{i,j}^{n+1}\cdot D_h\psi(u_{i,j}^{n+1})\right| \leq C\,\|\psi\|_\infty$$

for any bounded real function $\psi(r)$. In particular, if ψ is nondecreasing, this implies

$$h^2\sum_{n=0}^{N_T-1}\sum_{i,j}\Psi(u_{i,j}^{n+1})-\Psi(u_{i,j}^n) + \nu\,h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j} D_h u_{i,j}^{n+1}\cdot D_h\psi(u_{i,j}^{n+1}) \leq C\,\|\psi\|_\infty$$

where $\Psi(s) = \int_0^s \psi(r)dr$. Thus, since $|\Psi(s)| \leq \|\psi\|_\infty|s|$, one gets

$$\nu\,h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j} D_h u_{i,j}^{n+1}\cdot D_h\psi(u_{i,j}^{n+1}) \leq C\,\|\psi\|_\infty$$

where C only depends on the L^1-norm of the data. This is the desired a priori estimate; from which, using Fatou's lemma, we deduce

$$\int_Q |D\tilde{u}|^2\,\psi'(\tilde{u})\,dxdt \leq C\|\psi\|_\infty\,.$$

On account of the fact that u is bounded below, we can take for example $\psi(r) = 1 - e^{-\mu r}$ to deduce that $e^{-\mu\tilde{u}} \in L^2(0,T;H^1(\mathbb{T}^2))$ for any $\mu > 0$.

Thanks to (6.7), inequality (6.6) holds true not only for smooth functions φ but also for $\varphi \in H^1(Q)\cap L^\infty$, through a standard density argument. Moreover, there is no loss of generality in assuming that $u_0 \in H^1(\mathbb{T}^2)$, so we extend $\tilde{u}$ for negative t as identically equal to u_0. Then, we choose

$$\varphi(x,t) = \xi(t)\frac{1}{h}\int_{t-h}^t e^{\lambda T_k(\tilde{u})(x,s)}\,ds$$

where $T_k(r) = \min(r,k)$ and $\xi \in C_c^1[0,T)$. Using the monotone character of $s \mapsto e^{-\lambda s}$ we have (see Lemma 2.3 in [6])

$$\limsup_{h\to 0}\left\{\frac{1}{\lambda}\int_Q e^{-\lambda u}\varphi_t\,dxdt + \frac{1}{\lambda}\int_{\mathbb{T}^2} e^{-\lambda u_0}\,\varphi(0)\,dx\right\} \\ \leq -\int_Q \xi_t\int_0^u e^{-\lambda(r-T_k(r))}\,dr - \int_{\mathbb{T}^2}\xi(0)\int_0^{u_0} e^{-\lambda(r-T_k(r))}\,dr\,.$$

Moreover, $\frac{1}{h}\int_{t-h}^{t} e^{\lambda T_k(\tilde u)(x,s)}\, ds$ converges to $e^{\lambda T_k(u)}$ in $L^2(0,T;H^1(\mathbb{T}^2))$ and weak$-*$ in $L^\infty(Q)$, so we can pass to the limit as $h\to 0$ in the remaining terms of (6.6). Finally, we obtain

$$\begin{aligned}
&-\int_Q \xi_t \int_0^u e^{-\lambda(r-T_k(r))}\,dr - \int_{\mathbb{T}^2}\xi(0)\int_0^{u_0} e^{-\lambda(r-T_k(r))}\,dr\\
&+\nu\int_Q D\tilde u D\xi\, e^{-\lambda(\tilde u - T_k(\tilde u))}\,dxdt - \nu\,\lambda\int_{\{\tilde u>k\}} |D\tilde u|^2\, e^{-\lambda(\tilde u-T_k(\tilde u))}\,\xi\,dxdt\\
&+\int_Q H(x,D\tilde u)e^{-\lambda(\tilde u-T_k(\tilde u))}\,\xi\,dxdt \ge \int_Q F(\tilde m)e^{-\lambda(\tilde u-T_k(\tilde u))}\xi\,dxdt.
\end{aligned}$$

We conclude by letting $k\to\infty$, thanks to the dominated convergence theorem:

$$-\int_Q \tilde u\,\xi_t\,dxdt - \int_{\mathbb{T}^2} u_0\,\xi(0)\,dx + \nu\int_Q D\tilde u D\xi\,dxdt + \int_Q H(x,D\tilde u)\xi\,dxdt \ge \int_Q F(\tilde m)\,\xi\,dxdt\,,$$

for every $\xi\ge 0$. Since the reverse inequality was already obtained previously, in the end, we proved that u solves

$$-\int_Q \tilde u\,\xi_t\,dxdt - \int_{\mathbb{T}^2} u_0\,\xi(0)\,dx + \nu\int_Q D\tilde u D\xi\,dxdt + \int_Q H(x,D\tilde u)\xi\,dxdt = \int_Q F(\tilde m)\,\xi\,dxdt$$

for every $\xi\in C_c^1([0,T)), \xi\ge 0$, and therefore for every ξ. This concludes the proof that $\tilde u$ is a weak solution to the limit equation.

6.3. Stability of the discrete Fokker-Planck equation. We now pass to the limit in the discrete Fokker-Planck equation.

By (4.6), the L^1-compactness of $m_{h,\Delta t}$ and of $\widetilde{Du}_{h,\Delta t}$, we deduce the strong convergence in $L^1(Q)$ for the piecewise constant function which takes the value $m^n_{i,j}\nabla_q g(x_{i,j},[\nabla_h u^{n+1}]_{i,j})$ for (t,x) such that $|x_1-ih|<h/2$, $|x_2-jh|<h/2$, $t_n\le t<t_{n+1}$. Moreover, by the consistency assumption we have,

$$\begin{aligned}
h^2\Delta t\sum_{n=0}^{N_T-1}\sum_{i,j} m^n_{i,j}\nabla_q g(x_{i,j},[\nabla_h u^{n+1}]_{i,j})\cdot[\nabla_h\varphi]_{i,j} \to&\\
\to \int_Q \nabla_q g(x, D_1\tilde u, D_1\tilde u, D_2\tilde u, D_2\tilde u)\cdot(D_1\varphi, D_1\varphi, D_2\varphi, D_2\varphi)dxdt&\\
=\int_Q m\,\frac{\partial H}{\partial p}(x,D\tilde u)\cdot D\varphi\,dxdt&
\end{aligned}$$

Therefore, we can pass to the limit in the weak formulation and deduce that m is a weak solution of the Fokker-Planck equation.

We notice that the regularity $m[H_p(\cdot,D\tilde u)D\tilde u - H(\cdot,D\tilde u)]\in L^1(Q)$ follows from inequality (4.10), by using Fatou's lemma. Moreover, we also find that $m|H_p(\cdot,D\tilde u)|^2\in L^1(Q)$. The regularity $\tilde u,\tilde m\in C^0([0,T];L^1(\mathbb{T}^2))$ follows from properties of weak solutions, see [28].

Finally, this concludes the proof of Theorem 3.1.

Acknowledgements. The first author was partially funded by the ANR projects ANR-12-MONU-0013 and ANR-12-BS01-0008-01.

REFERENCES

[1] Y. Achdou. Finite difference methods for mean field games. In P. Loreti and N. A. Tchou, editors, *Hamilton-Jacobi equations: approximations, numerical analysis and applications*, volume 2074 of *Lecture Notes in Math.*, pages 1–47. Springer, Heidelberg, 2013.

[2] Y. Achdou, F.J. Buera, J-M. Lasry, P-L. Lions, and B. Moll. Partial differential equation models in macroeconomics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2028), 2014.

[3] Y. Achdou, F. Camilli, and I. Capuzzo-Dolcetta. Mean field games: numerical methods for the planning problem. *SIAM J. Control Optim.*, 50(1):77–109, 2012.

[4] Y. Achdou, F. Camilli, and I. Capuzzo-Dolcetta. Mean field games: convergence of a finite difference method. *SIAM J. Numer. Anal.*, 51(5):2585–2612, 2013.

[5] Y. Achdou and I. Capuzzo-Dolcetta. Mean field games: numerical methods. *SIAM J. Numer. Anal.*, 48(3):1136–1162, 2010.

[6] D. Blanchard and A. Porretta. Stefan problems with nonlinear diffusion and convection. *J. Differential Equations*, 210(2):383–428, 2005.

[7] L. Boccardo and T. Gallouët. Nonlinear elliptic and parabolic equations involving measure data. *J. Funct. Anal.*, 87(1):149–169, 1989.

[8] L. Boccardo, F. Murat, and J.-P. Puel. Existence de solutions non bornées pour certaines équations quasi-linéaires. *Portugal. Math.*, 41(1-4):507–534 (1984), 1982.

[9] P. Cardaliaguet. Notes on mean field games. preprint, 2011.

[10] E. Carlini and F. J. Silva. A semi-lagrangian scheme for a degenerate second order mean field game system. Technical report.

[11] E. Carlini and F. J. Silva. A fully discrete semi-Lagrangian scheme for a first order mean field game problem. *SIAM J. Numer. Anal.*, 52(1):45–67, 2014.

[12] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In *Handbook of numerical analysis, Vol. VII*, Handb. Numer. Anal., VII, pages 713–1020. North-Holland, Amsterdam, 2000.

[13] R. Eymard, T. Gallouët, D. Hilhorst, and Y. Naït Slimane. Finite volumes and nonlinear diffusion equations. *RAIRO Modél. Math. Anal. Numér.*, 32(6):747–761, 1998.

[14] T. Gallouët, A. Larcher, and J. C. Latché. Convergence of a finite volume scheme for the convection-diffusion equation with L^1 data. *Math. Comp.*, 81(279):1429–1454, 2012.

[15] D. A. Gomes, J. Mohr, and R. R. Souza. Discrete time, finite state space mean field games. *J. Math. Pures Appl. (9)*, 93(3):308–328, 2010.

[16] D. A. Gomes and J. Saúde. Mean field games models—a brief survey. *Dyn. Games Appl.*, 4(2):110–154, 2014.

[17] O. Guéant. Mean field games equations with quadratic Hamiltonian: a specific approach. *Math. Models Methods Appl. Sci.*, 22(9):1250022, 37, 2012.

[18] O. Guéant. New numerical methods for mean field games with quadratic costs. *Netw. Heterog. Media*, 7(2):315–336, 2012.

[19] O. Guéant, J-M. Lasry, and P-L. Lions. Mean field games and applications. In *Paris-Princeton Lectures on Mathematical Finance 2010*, volume 2003 of *Lecture Notes in Math.*, pages 205–266. Springer, Berlin, 2011.

[20] M. Huang, P.E. Caines, and R. P. Malhamé. Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized ϵ-Nash equilibria. *IEEE Trans. Automat. Control*, 52(9):1560–1571, 2007.

[21] A. Lachapelle, J. Salomon, and G. Turinici. Computation of mean field equilibria in economics. *Math. Models Methods Appl. Sci.*, 20(4):567–588, 2010.

[22] J-M. Lasry and P-L. Lions. Jeux à champ moyen. I. Le cas stationnaire. *C. R. Math. Acad. Sci. Paris*, 343(9):619–625, 2006.

[23] J-M. Lasry and P-L. Lions. Jeux à champ moyen. II. Horizon fini et contrôle optimal. *C. R. Math. Acad. Sci. Paris*, 343(10):679–684, 2006.

[24] J-M. Lasry and P-L. Lions. Mean field games. *Jpn. J. Math.*, 2(1):229–260, 2007.

[25] P-L. Lions. Cours du Collège de France. http://www.college-defrance.fr/default/EN/all/equ_der/, 2007-2011.

[26] A. Porretta. On the planning problem for a class of mean field games. *C. R. Math. Acad. Sci. Paris*, 351(11-12):457–462, 2013.

[27] A. Porretta. On the planning problem for the mean field games system. *Dyn. Games Appl.*, 4(2):231–256, 2014.

[28] A. Porretta. Weak Solutions to Fokker–Planck Equations and Mean Field Games. *Arch. Ration. Mech. Anal.*, 216(1):1–62, 2015.

[29] J. Simon. Compact sets in the space $L^p(0,T;B)$. *Ann. Mat. Pura Appl. (4)*, 146:65–96, 1987.